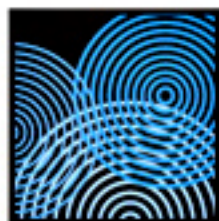


VISIT...

LANZAROTE
Caliente.COM



CAMBRIDGE
STUDIES IN
PHILOSOPHY



MORALITY IN A
NATURAL WORLD

SELECTED ESSAYS
IN METAETHICS

DAVID COPP

This page intentionally left blank

Morality in a Natural World

The central philosophical challenge of metaethics is to account for the normativity of moral judgment without abandoning or seriously compromising moral realism. In *Morality in a Natural World*, David Copp defends a version of naturalistic moral realism and argues that it can accommodate the normativity of morality. Largely because of the difficulty in accounting for normativity, naturalistic moral realism is often thought to face special metaphysical, epistemological, and semantic problems. In the ten essays included in this volume, Copp defends solutions to these problems. Three of the essays are new, while seven have previously been published. All of them are concerned with the viability of naturalistic and realistic accounts of the nature of morality or, more generally, with the viability of naturalistic and realistic accounts of reasons.

David Copp is professor of philosophy at the University of Florida. He is the author of *Morality, Normativity and Society* and has edited and co-edited several volumes, including *The Oxford Handbook of Ethical Theory*. He served for many years as an editor of the *Canadian Journal of Philosophy* and is currently an associate editor of *Ethics* and the subject editor for metaethics of the *Stanford Encyclopedia of Philosophy*.

CAMBRIDGE STUDIES IN PHILOSOPHY

General Editors

Jonathan Lowe (University of Durham)
Walter Sinnott-Armstrong (Dartmouth College)

Advisory Editors

Jonathan Dancy (University of Texas, Austin)
John Haldane (University of St. Andrews)
Gilbert Harman (Princeton University)
Frank Jackson (Australian National University)
William G. Lycan (University of North Carolina at Chapel Hill)
Sydney Shoemaker (Cornell University)
Judith J. Thomson (Massachusetts Institute of Technology)

Recent Titles

David Lewis *Papers on Metaphysics and Epistemology*
Raymond Martin *Self-Concern*
Annette Barnes *Seeing Through Self-Deception*
Michael Bratman *Faces of Intention*
Amie Thomasson *Fiction and Metaphysics*
David Lewis *Papers on Ethics and Social Philosophy*
Fred Dretske *Perception, Knowledge, and Belief*
Lynne Rudder Baker *Persons and Bodies*
John Greco *Putting Skeptics in Their Place*
Ruth Garrett Millikan *On Clear and Confused Ideas*
Derk Pereboom *Living Without Free Will*
Brian Ellis *Scientific Essentialism*
Alan H. Goldman *Practical Rules: When We Need Them and When We Don't*
Christopher Hill *Thought and World*
Andrew Newman *The Correspondence Theory of Truth*
Ishtiyaque Haji *Deontic Morality and Control*
Wayne A. Davis *Meaning, Expression and Thought*
Peter Railton *Facts, Values, and Norms*
Rosanna Keefe *Theories of Vagueness*
David Armstrong *Truth and Truthmakers*
Keith Frankish *Mind and Supermind*
Joshua Gert *Brute Rationality: Normativity and Human Action*
Jane Heal *Mind, Reason and Imagination*
Jonathan Kvanvig *The Value of Knowledge and the Pursuit of Understanding*
Andrew Melnyk *A Physicalist Manifesto*
William S. Robinson *Understanding Phenomenal Consciousness*
Noah Lemos *Common Sense*
Michael Smith *Ethics and the A Priori*
Folke Tersman *Moral Disagreement*
Alexander R. Pruss *The Principle of Sufficient Reason*
Joseph Mendola *Goodness and Justice*

Morality in a Natural World

SELECTED ESSAYS IN METAETHICS

DAVID COPP

University of Florida



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521863711

© David Copp 2007

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2007

ISBN-13 978-0-511-29522-5 eBook (EBL)

ISBN-10 0-511-29522-7 eBook (EBL)

ISBN-13 978-0-521-86371-1 hardback

ISBN-10 0-521-86371-6 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Marina

Contents

<i>Preface</i>	<i>page xi</i>
<i>Acknowledgments</i>	<i>xiii</i>
Introduction	1
Part One. Naturalism: Epistemology and Metaphysics	
1. Why Naturalism?	33
2. Four Epistemological Challenges to Ethical Naturalism: Naturalized Epistemology and the First-Person Perspective	55
3. Moral Naturalism and Self-Evident Moral Truths	93
4. Moral Necessities in a Contingent World	113
Part Two. Referring to Moral Properties	
5. Realist-Expressivism: A Neglected Option for Moral Realism	153
6. Milk, Honey, and the Good Life on Moral Twin Earth	203
7. Referring to Moral Properties: Moral Twin Earth, Again	230
Part Three. Naturalism and Normativity	
8. Moral Naturalism and Three Grades of Normativity	249
9. The Ring of Gyges: Overridingness and the Unity of Reason	284
10. The Normativity of Self-Grounded Reason	309
<i>Index</i>	<i>355</i>

Preface

This volume brings together ten essays in metaethics that I have written over the past decade. Three are previously unpublished. All of them aim in one way or another to defend the viability of a naturalistic and realistic account of the nature of morality. They discuss problems for naturalism, chiefly the problem of explaining the normativity of moral judgment, and they suggest or defend solutions to the problems.

The point of reprinting the articles is that, taken together, and with the addition of the three new essays, they develop a systematic defense of moral naturalism. Moreover, some of them initially appeared in out-of-the-way places. I see difficulties in each of them, certainly in the previously published essays, difficulties that I wish I had noticed much earlier. I have largely resisted the temptation to make substantive changes, however, because some people will have read the original versions of the essays and I did not want to cause confusion about my views. For this reason, the seven previously published essays in the book are reproduced largely without alteration, except for minor changes. I have changed the style of the notes, and I have added a few substantive notes. Because of this, the notes have been renumbered in some cases. When I wrote the essays, I intended them to be read individually, which means that some points are repeated in more than one, but the result is that each of the chapters in the book can be understood without reading any of the others. The introduction aims to put the chapters into context and to explain some ideas that lie in the background of my arguments.

During the past ten years, I have been fortunate in being a member of the philosophy departments at the University of California, Davis; Bowling Green State University; and the University of Florida. Each of these universities generously gave me time for research. I also enjoyed

very welcome fellowships with the Philosophy Program at the Research School of Social Sciences, Australian National University; the Center for Applied Ethics, University of British Columbia; and the Social Philosophy and Policy Center, Bowling Green State University. I would like to thank each of these institutions, and especially, of course, the people who work in them, for their valuable assistance.

So many people have given me help in developing my ideas that I cannot hope to remember them all. In each of the essays I thank by name the people I can remember who gave me comments and suggestions, and I thank the audiences that heard me lecture on the topics of the essays. I am enormously grateful for the time and effort that all of these people invested in helping me.

There are some colleagues and friends to whom I owe special thanks, both for their stimulation and intellectual help and for their friendly encouragement. I would especially like to mention a few colleagues at Davis, Bowling Green, and Florida who have had an especially important impact on my thinking, namely, Jerry Dworkin, Michael Jubien, Jeff King, David Sobel, and Jon Tresan. I was very lucky to have them as colleagues. For delightful collegial discussions of issues in moral philosophy, I would like to thank the Davis Ethics Discussion Group, the Ohio Reading Group in Ethics, and the Gator Philosophy and Ethics Discussion Group at the University of Florida. Michael Ridge gave me extensive comments on several of the essays included in this book as well as on my proposal to Cambridge University Press. Walter Sinnott-Armstrong invited me to put together this collection for Cambridge, and he gave me valuable feedback on many of the chapters, including the introduction. He has encouraged me in the development of my views ever since we first talked about them. I owe him and the others I have mentioned a very large debt of gratitude.

Marina Oshana has made life easy and pleasant for me and has helped me on many occasions to clarify my thinking with her comments on essays included here. Five cats have shared our home over the years and they have kept me awake to the rhythm of life outside my study. Without such good fortune at home, I could not have written these essays.

Acknowledgments

The essays published here as chapters 3, 4, and 7 have not previously been published. The remaining seven essays originally appeared in the journals and volumes listed here below. I am very grateful to the publishers for giving their permission for this reprinting. Chapter 1, “Why Naturalism?” originally appeared in *Ethical Theory and Moral Practice* 6 (2003): 179–200, and is reprinted with the kind permission of Springer Science and Business Media. Chapter 2, “Four Epistemological Challenges to Ethical Naturalism: Naturalized Epistemology and the First-Person Perspective,” originally appeared in *Canadian Journal of Philosophy* supp. vol. 26 (2001): 31–74. Chapter 5, “Realist-Expressivism: A Neglected Option for Moral Realism,” originally appeared in *Social Philosophy and Policy* 18 (2001): 1–43. Chapter 6, “Milk, Honey, and the Good Life on Moral Twin Earth,” originally appeared in *Synthèse* 124 (2000): 113–137, and is reprinted with the kind permission of Springer Science and Business Media. Chapter 8, “Moral Naturalism and Three Grades of Normativity,” originally appeared in Peter Schaber, ed., *Normativity and Naturalism* (Frankfurt: Ontos-Verlag, 2004), pp. 7–45. Chapter 9, “The Ring of Gyges: Overridingness and the Unity of Reason,” originally appeared in *Social Philosophy and Policy* 14 (1997): 86–106. Finally, chapter 10, “The Normativity of Self-Grounded Reason,” originally appeared in *Social Philosophy and Policy* 22 (2005): 165–203.

Introduction

Our thoughts about our moral thinking are Janus-faced. On the one hand, we intuitively and pre-theoretically think as moral ‘realists’ – we take our moral convictions to be beliefs in just the way that our convictions about the weather are beliefs, and of course we take our convictions to be true. Indeed, we take some of them to be self-evidently true. On the other hand, we find ourselves facing intuitively significant challenges that can make moral realism seem problematic or even completely implausible. Ordinary reflection tells us that our moral convictions are different in nature from most other beliefs, such as our beliefs about the weather. Moral judgments are directly relevant to decisions and choices in a way that differs from the way that beliefs about the weather might be relevant to decisions and choices. Intuitively, moreover, a moral judgment speaks to what ‘ought to be the case’ rather than to what ‘is the case.’ We can introduce a term to talk about this. We can say that, unlike judgments about the likelihood of rainfall or the like, moral judgments are ‘normative.’¹ Unfortunately, however, it can easily seem dubious that there could be something

I am grateful to Marina Oshana and Walter Sinnott-Armstrong for helpful comments on an earlier draft of this introduction.

- 1 I do not know of a better word for the phenomenon at issue than “normativity.” I want to avoid the term “prescriptivity” because it has been given a technical meaning in Hare 1952. I want to avoid the term “action-guiding” because the thesis that moral judgments are ‘action-guiding’ tends to be associated with the thesis, often called “judgment internalism,” that there is an ‘internal’ connection between moral belief and appropriate motivation. I do not want to use a terminology that suggests that the normativity of moral judgment is simply a matter of the truth of judgment internalism. See below, in this introduction, and chapter 8. See also Copp 1995b.

in the world *as it is* that makes true a judgment about what *ought* to be the case. There is an obvious tension between these two sides to our thinking.

The chief philosophical challenge facing ‘metaethical’ theory – the theory of the nature and truth conditions of moral judgment – is to account for the normativity of moral judgment without abandoning or seriously compromising moral realism. There are, of course, two ways to attempt to avoid the challenge.

First, one might deny that there is such a thing as normativity, or one might abandon the goal of explaining normativity. I think, however, that it is beyond question that moral judgment is normative, although there is room for disagreement about what normativity comes to. Moreover, I think it is beyond question that moral philosophy must aim to explain the central features of moral thought and discourse. Hence I think that an adequate metaethical theory must explain what the normativity of moral judgment consists in. I call this the ‘normativity constraint.’ Theories that simply postulate primitive unexplained *sui generis* normative moral properties or that help themselves to an unexplained normative notion of reasonableness or rationality are not satisfying. They leave a mystery at the foundation of our moral thinking.

Second, one might abandon or compromise moral realism. ‘Noncognitivism’ takes moral conviction to be a kind of conative state rather than strictly speaking a state of belief, while ‘nihilism’ or the ‘error theory’ denies that any of our basic moral convictions are strictly speaking true. Moral realism is, however, the ‘default view,’ or so I will argue. Indeed, I believe that ‘moral naturalism’ is the default view – taking moral naturalism to be the combination of moral realism with naturalism.² To be more exact, moral naturalism is the position, roughly, and in part, that our moral beliefs ascribe moral characteristics to things, characteristics such as goodness and rightness, and that these characteristics are *natural* characteristics, relevantly similar to ordinary properties of things, such as meteorological or economic properties. Moral naturalism is not beyond question; a successful argument that it cannot accommodate the normativity of moral thought should lead us to abandon it. I will argue, however, that it is the default view.

2 Noncognitivism and the error theory can be counted as forms of naturalism in a broad sense since they are compatible with the view that all facts are natural facts.

My goal in this book is to develop and defend a kind of moral naturalism and to argue that it can explain the normativity of moral judgment without compromising moral realism. It can capture both sides of our intuitive view. I have presented such a theory before, in my book *Morality, Normativity, and Society*.³ The chapters in the present volume build on the ‘society-centered theory’ that I proposed in that first book. My fundamental goal here, however, is to support the viability of moral naturalism.

The chief purpose of this introduction is to explain more fully the normativity constraint, as well as moral naturalism, and to introduce the society-centered theory. A second purpose is to explain how the issues discussed in individual chapters of the book are related to the defense of moral naturalism. In section 1, I explain the normativity constraint. In section 2, I explain why I believe that moral realism is the default view. In section 3, I explain why I believe that naturalism is also a default view. In section 4, I introduce the society-centered theory. Of course, there are many questions about it that I cannot address here. One important distinction that I need to explain is between the ‘constructivist’ version of the theory that I presented in my first book and the ‘nonconstructivist’ version that is at work in the present book. I believe that the nonconstructivist version is preferable. In section 5, I provide an overview of the book.

1. THE NORMATIVITY CONSTRAINT

The normativity constraint says that *an adequate metaethical theory must explain what the normativity of moral judgment consists in*.

Compare the propositions that I morally *ought* to give to famine relief, or that it would be *good* of me to do so, with the proposition that I *rarely* give to famine relief. The latter, nonmoral, claim is simply descriptive of an aspect of my behavior, but the moral claims are not merely descriptive. They are prescriptive or evaluative, and they are prescriptive or evaluative in virtue of what they say, or in virtue of their content. They are normative, and because of this, my belief that I ought to give to famine relief, or that it would be good of me to do so, has a direct and immediate relevance to decisions or choices I might make – a relevance of a kind that a belief that I rarely give to famine relief does not have. Moral beliefs in general have a

3 Copp 1995a.

characteristic kind of direct relevance to decisions or choices because the propositions that are their objects are normative.

From the perspective of a moral realist, as I will explain, the normative proposition that honesty is good differs from the nonnormative proposition that honesty is rare only in that they ascribe different properties. The difference between them must therefore lie in the nature of the properties involved. To explain the fact that the proposition that honesty is good is normative, while the proposition that honesty is rare is not, we must take it that the property of being good is normative, while the property of being rare is not normative. For similar reasons, we need to see other moral properties as normative. An adequate realist theory would need to explain what this normativity consists in.

Moral properties, if any exist, are necessarily normative; a property would not count as a moral property *unless* it were (in some way) normative. I call this idea ‘normative internalism,’ and if it is correct, it rules out a familiar kind of moral naturalism that has been proposed by a number of philosophers, including Richard Boyd, David Brink, Peter Railton, and Nicholas Sturgeon.⁴ The position they share is commonly known as “Cornell moral realism” because of the influence of Cornell philosophers in defending it. According to Cornell realism, the normativity of a moral property is ‘external’ to it – it is not essential to it. It is a matter of how the property happens to be related to our motivational states. People typically are motivated to avoid wrongdoing, for instance, because of what wrongdoing involves in the treatment of people. But it is a contingent matter that people are motivated in this way, and so, on the Cornell position, it is a contingent matter that moral properties are normative. Moreover, it appears that the Cornell view would implausibly count sweetness as a normative property, since people are typically motivated to seek sweet things. I believe, then, that Cornell realism does not provide an adequate account of the normativity of moral properties. This failure undermines its defense of moral naturalism, for to show that a natural property could be a moral property, we need to show that a natural property could be normative, and to show this, we need an account of what its normativity would consist in.⁵

4 Boyd 1988; Brink 1989; Railton 1986; Sturgeon 1984.

5 I argued this point in Copp 1990. Stephen Darwall, Allan Gibbard, and Peter Railton accept my claim that in order to show that a property is a moral property, one must show that it is normative (1992, 128 n. 30). Alexander Miller claims, however, that my arguments merely show that Cornell realism is compatible with a kind of externalism according to which a person might believe that an action would be morally wrong and yet coherently deny that

Cornell realism leaves open the possibility of believing that one morally ought to do something without being motivated in the least to do it. I agree with Cornell realism about this. But the contrary view, which is standardly called “judgment internalism,”⁶ is widely accepted. According to judgment internalism, it is a conceptual truth, and necessarily the case, that if a person judges he or she ought morally to do something, he or she is motivated to some degree to do it. It might seem that this doctrine accounts for the normativity of moral judgment. I believe, however, that judgment internalism is false.

There are familiar arguments against judgment internalism. It appears, for example, that people who are depressed might lack any motivation to do what they believe they morally ought to do, and people with unusual second-order beliefs about morality might also lack appropriate motivation. I once presented the following putative counter-example to judgment internalism, the case of Alice:⁷

Alice was raised to believe . . . that our moral obligations are determined by the commands of God. She was also raised to believe that God is a vengeful ruler and that He wills us to take an eye for an eye. On the principle of an eye for an eye, Alice believes that capital punishment is obligatory in cases of murder, and she believes she has an obligation to support capital punishment. But she is deeply compassionate, and she is quite out of sympathy with what she takes to be God’s vengefulness. Because of her compassion she is not motivated in the least to support capital punishment. She is in fact active in opposing it, even though she believes she is morally forbidden to do so.

This case does not seem to be ruled out on conceptual grounds. Or consider the case of Huckleberry Finn. Huck believes he is morally obligated to turn his friend Jim over to the authorities because Jim is an escaped slave. But Huck does not turn him in, and it seems coherent to suppose that Huck is not motivated in the least to do so.⁸ Given these examples, and other examples that are similar in nature, I conclude that judgment internalism is false.

he or she has a reason not to do it (Miller 2003, 160–162). But this is not my objection. My objection is that Cornell realism fails to show the existence of any normative properties, and since moral properties are necessarily normative, it fails to show that there are any moral properties.

6 The terminology is from Darwall 1983, 54–55. Brink calls the position “belief internalism” (1989, 40). In chapter 8 of this book, I call it “motivational internalism.”

7 Copp 1995b, 190–191.

8 Ibid., 204.

Moreover, for reasons I explain in chapter 8, I think that even if judgment internalism were true, the existence of a conceptual link between moral belief and motivation would not be adequate to account for the normativity of moral judgment. This is important since, I believe, people resist the counter-arguments to judgment internalism mainly because they do not see how they could otherwise account for the normativity of moral judgment. Indeed, I think there is a tendency to confuse judgment internalism with the different idea, which I believe to be true, that it is a necessary truth that moral belief is normative. My own view is that *normativity* is internal to moral judgment although *motivation* is external to it. If I am correct, we need a new strategy for explaining normativity.

A fully satisfying account of the normativity of moral judgment must explain the link between moral belief and decision. Moral belief has a characteristic kind of direct relevance to decisions, which needs to be explained, and morality may seem to have a kind of final authority over our decisions and actions, which would also need to be explained. I tackle these issues in chapters 8 through 10. In chapter 8, I systematically explore the difficulty of accounting for the normativity of morality, and I argue that a well-designed naturalistic theory can meet the challenge. In chapter 9, I argue that moral considerations do not have the automatic kind of ‘overriding’ authority over our decisions that people sometimes think they do. In chapter 10, I explain that moral beliefs that flow from our values do have an immediate and direct relevance to rational decision making.

The normativity constraint has powerful implications for moral theory. The constraint rules out, or at least deems to be inadequate, realist theories that fail to explain the normativity of moral properties. It implies that nonnaturalistic theories that postulate *sui generis* unexplained normative moral properties are inadequate. It also rules out versions of moral naturalism that fail to explain normativity.

2. THE DEFAULT VIEW: MORAL REALISM

When I say that moral realism is the ‘default view,’ I mean it is the view about moral judgment that one is naturally led to if one approaches the subject without prior theoretical commitments. I think it is natural to hold that our moral ‘convictions’ are *beliefs* in just the way that beliefs about the weather are beliefs – although, obviously, they have a different subject matter. Moreover, it is natural to think that at least some of our moral beliefs are true. And it is natural to think that our moral beliefs ascribe

moral ‘properties’; for example, it is natural to think that the belief that lying is wrong represents lying as having a certain characteristic or as ‘being a certain way.’⁹

As I explain briefly in chapters 5 and 8, we can usefully distinguish five doctrines that are included in the realist’s position:¹⁰

- (1) There are moral properties (and relations).¹¹ There is, for example, such a thing as wrongness.
- (2) Some moral properties are instantiated. For example, some actions are wrong.
- (3) Moral predicates are used to ascribe moral properties. When we call an action “wrong,” we are ascribing to it the property wrongness.
- (4) Moral assertions express moral beliefs. When we call an action “wrong,” we are expressing the belief that the action is wrong.
- (5) Moral properties, in being properties, have the metaphysical status that any other property has, whatever that status is.¹²

Given the complexity of this characterization, one might doubt that realism can be the default position. But the first four doctrines were implicit in my initial intuitive sketch, and the fifth doctrine is intended merely to express the idea that the moral characteristics of things are, quite simply, properties.

The reason we need this fifth doctrine, as I explain in chapter 5, is to distinguish moral realism from a kind of sophisticated noncognitivist antirealism that accepts ‘deflationary’ versions of the first four doctrines. I

9 Moral realism has been called “descriptivism,” but this is not an apt label, for a realist should deny that moral claims are merely descriptive. Moral properties are normative, which means that moral propositions do not merely describe. They also evaluate, or proscribe, or the like. For instance, the proposition that torture is wrong ‘describes’ torture but also evaluates it.

10 I explain these doctrines in Copp 2006a, 6–8.

11 In what follows, I treat relations, such as the relation of being better than, as a kind of property.

12 That is, clause (5) says, clause (1) is to be interpreted such that the term “property,” as it occurs there, ascribes the same metaphysical status to moral properties, such as *wrongness*, as it ascribes to a nonmoral property such as *redness* when it is predicated of such a property. Moral realism is compatible with any theory that acknowledges the existence of properties or ‘characteristics’, or ‘ways that things are,’ including nominalism. The moral realist says that moral properties have the metaphysical status that any other property has, *whatever* that is. Some philosophers would deny that there are any properties at all. But I take it that they do not mean to deny that red things have the ‘characteristic’ of being red. They mean to reject the standard philosophical theories about the nature of such characteristics. If they would agree that sentences such as “There is such a thing as redness” can be used to express truths, they may be in a position to accept moral realism. I am grateful to Thomas Hofweber and Michael Jubien for helpful discussions about the nature of properties.

have in mind Simon Blackburn's "quasi-realism," for example. Blackburn agrees that everyday moral discourse has a "realist surface."¹³ He allows that there is no objection to our ordinary practice of speaking of ourselves as having true moral beliefs, nor to speaking of wrongness as a property. But he holds that a plausible metaphysics would not postulate moral properties and that it would deny that the states of mind that we call "moral beliefs" are cognitive states that 'represent' things as having moral properties. They are not beliefs, strictly speaking. The fifth doctrine distinguishes moral realism from Blackburn's view. It says in effect that an adequate metaphysics would give the same account of the status of moral properties as it gives of the metaphysical status of nonmoral properties such as meteorological properties. This leaves it open what this status might be.

The core idea of noncognitivism is that the state of mind expressed by a person in making a basic moral claim is not, properly speaking, a belief or any other kind of cognitive state but is, instead, a *conative* state or a *motivational* state, akin to a desire. A fully developed version of noncognitivism would need to say exactly what kind of state of mind is involved, but for convenience, we may say that it is an 'attitude.' Blackburn speaks of "stances."¹⁴ Using this terminology, a noncognitivist might say that a person who 'thinks' that torture is wrong therein has an attitude of disapproval toward torture rather than a belief that 'represents' torture as being a certain way. This is difficult to accept. When, in thinking, I move from the thought that torture is widespread to the 'thought' that torture is morally appalling, there is a shift in the content of my thoughts, and perhaps also in the feelings that accompany them, but I do not notice a shift in their *nature*, from cognitive to conative. The one seems to be a thought just in the way that the other is a thought.¹⁵

The most familiar argument in favor of noncognitivism is an argument from judgment internalism. Noncognitivists typically take judgment internalism to support the proposition that moral judgments are motivational states, akin to desires and other conative states.¹⁶ As I have said, however, I believe that judgment internalism is false.

It is important to recognize, nevertheless, that moral realism is compatible with the view that moral assertions express conative states of mind.

13 The phrase "realist surface" and the term "quasi-realism" are used in Blackburn 2006.

14 For an overview of the position, see *ibid.*

15 There are important technical objections to noncognitivism. For one thing, to account for the workings of moral language, noncognitivism is forced to add complexity to its semantics of a kind that would be avoided on a realist theory. See Copp 1995a, 15–19.

16 See Blackburn 2006, 149–150.

Moral realists hold that moral assertions express beliefs – cognitive states that have representational content – but they need not deny that moral assertions might also express conative states such as approval or disapproval. I explore this idea in chapter 5. There are various familiar pejorative and commendatory predicates that, in standard and literal usage, both ascribe properties and express attitudes. Frege called these predicates “colored.”¹⁷ Realist-expressivism is the view that moral predicates, such as “right,” “wrong,” “good,” and “bad,” are colored terms that, in standard and literal usage, are used both to ascribe a moral property and to express an appropriate attitude. If this is correct, I believe it can explain certain intuitions that lead people to think that a person who has a moral belief must have an appropriate corresponding conative attitude of some kind.

Noncognitivist expressivism is not the only alternative to moral realism. There is also the “error theory” of J. L. Mackie. According to the error theory there are no moral properties; moreover, because of this, all basic moral propositions are false.¹⁸ Mackie’s theory entails, for example, that it is false that lying is wrong.¹⁹ Indeed, it follows from the error theory that nothing is morally wrong, not even torture. But this is very difficult to believe.

Mackie’s most interesting argument for the error theory is the so-called argument from queerness, which turns, in effect, on the claim that no natural property could be normative, that a normative property would be metaphysically queer. This argument is important, but I believe it is unsuccessful. My answer to it is found in chapter 8.

There are problems, then, with both of the antirealist alternatives to moral realism. Given this, and given that moral realism is the default position, I focus on developing and articulating a realist position.

17 Frege 1984c, 161; 1984b, 185; 1984d, 357.

18 ‘Basic’ moral propositions are, I stipulate, propositions that ascribe moral properties to things. The proposition that torture is wrong is basic. The proposition that either torture is wrong or torture is widespread is not basic, and nor is the proposition that it is not the case that torture is wrong. In Copp 1995a, I call basic moral propositions “paradigmatic.”

19 Mackie 1977, ch. 1. There are problems in the interpretation of the theory. On certain views, if there is no property ascribed by “wrong,” then sentences such as “Lying is wrong” would not even express propositions. On these theories, the error theory would commit Mackie to viewing such sentences as meaningless. However, Mackie himself takes the theory to imply that basic moral claims, such as that lying is wrong, are false, not that sentences such as “Lying is wrong” fail to express propositions and are meaningless. I assume that a plausible semantics would provide a way of understanding Mackie’s theory that would avoid this problem. An alternative reading of the theory might take it to say that wrongness is a property that could not possibly be instantiated. This does not seem to have been Mackie’s view, however.

3. THE DEFAULT VIEW: MORAL NATURALISM

Moral naturalism accepts the five doctrines that are characteristic of moral realism and adds the following:

(6) Moral properties are natural properties.

Although the meaning of this doctrine is controversial, the idea is that moral properties are ‘natural’ in the way that, say, the property of being irascible, the property of being a hurricane, and the property of being inflationary, are ‘natural.’ They are ordinary and unexceptional. Moral properties obviously differ in important ways from psychological, meteorological, and economic properties. Most importantly, they are normative. But the naturalist wants to say that at a fundamental epistemological and metaphysical level, their status is no different from the status of these other properties.

The naturalist’s view is, I believe, the default view; it is the view that one would intuitively be led to if one approached the subject without prior theoretical commitments. I say this because, pre-philosophically, it seems obvious that one can run up against moral goodness and badness (and so on) in the natural world, just as one can run up against inflationary conditions or hurricanes. Pre-philosophically, it seems obvious that one can come face to face with the badness of people and with goodness in people, just as one can come face to face with a person’s irascibility. Badness in a person’s character, for instance, could lead him or her to do wrong. It could lead to cruelty. Goodness could lead a person to do right by others. In this way the moral properties of people can play a psychological role in shaping their actions.²⁰ For this reason and others it seems plausible that these properties are of a piece with psychological properties, such as the property of being irascible – that is, they are equally natural. Philosophical arguments might lead one to think that this is not so, but before we consider such arguments, I think the naturalist’s position will seem difficult to deny.

To be sure, it is not clear how best to distinguish between natural and nonnatural properties. Intuitively, the natural world is the world around us, the world that we know about and are in contact with by means of the senses. In chapter 1, I propose that, for the purpose of explicating moral naturalism, we should take natural properties to be *empirical* properties. That is, a natural property is such that any substantive knowledge we

²⁰ Sturgeon 1984.

have about its instantiation is empirical. Or better, a property is a natural property just in case any synthetic proposition about its instantiation is *empirically defeasible* in that there can be empirical evidence that it is false.²¹

In chapter 1, I also present an argument from disagreement to support the idea that moral properties are natural properties in the sense explained. I argue that evidence that others disagree with a belief that *p*, and disagree persistently, would be empirical evidence against *p* – provided that *p* is synthetic.²² Synthetic moral propositions are therefore empirically defeasible, and so moral properties qualify as natural properties. In chapter 3, however, I express some doubts about this argument.

If I am correct that moral naturalism is the default position, moral naturalists do not need to argue for their view. Of course they do have the burdens of explaining it and of dealing with many serious challenges to it. But if moral naturalism is the default position, those who think it is false must shoulder the burden of presenting arguments against it.

The most influential argument against naturalism is the famous “open question argument,” which was proposed by G. E. Moore.²³ In the form in which it is discussed nowadays, and ignoring various niceties, the argument is as follows:

1. Any version of moral naturalism is committed to an ‘identity thesis’ of the form $\langle \text{Mness is identical to Nness} \rangle$, where Mness is a moral property and Nness is a natural property.
2. If such an identity thesis were true, it would be a conceptual truth.

21 This is intended as a proposal about how best to explicate the idea of a natural property. It is an ‘explicating proposal.’ Moral naturalism combines moral realism with the doctrine that moral properties are natural, and an explicating proposal adds that a property is natural just in case it is *N*, for some appropriate *N*. Let EP be such a proposal. EP will not be plausible unless the thesis that moral properties are *N* is entailed by, or is at least compatible with, most positions that would standardly be classified as versions of moral naturalism. Despite this, however, EP does not commit us to holding that a philosopher’s theory is a version of moral naturalism if and only if it entails that moral properties are *N*. It is not a proposal about how to classify philosophers’ positions. The thesis that moral properties are *N* is, by the lights of EP, the thesis that best explicates the idea that moral properties are natural, but the fundamental idea that all moral naturalists share is that moral properties are natural, not that moral properties are *N*. It is likely that some naturalists will reject EP. I am grateful to Greg Ray and Gene Witmer for helpful discussion of this point.

22 If *p* is a conceptual truth, then others’ persistent disagreement with our belief that *p* is not evidence that *p* is false. To be sure, such disagreement might mean that we are less justified in our belief than we otherwise would be, and it might be evidence that *p* is not a conceptual truth.

23 Moore 1903, sec. 13, pp. 67–68.

3. If such a thesis were a conceptual truth, it would not be an open question whether something that is N is also M.
4. For any natural property, N_{ness}, and any moral property, M_{ness}, it is an open question whether something that is N is also M.
5. Hence no such identity thesis is true.
6. Hence, moral naturalism is false.

I discuss this argument very briefly in chapter 4. I believe that premises 2 and 3 are implausible, as Michael Smith, Nicholas Sturgeon, and others have argued.²⁴

A second influential argument against naturalism is the ‘is/ought argument.’ According to this argument,

1. There are no cogent inferences from nonmoral premises to moral conclusions.
2. Any version of moral naturalism is committed to an ‘identity thesis’ of the form (M_{ness} is identical to N_{ness}), where M_{ness} is a moral property and N_{ness} is a natural property.
3. If such an identity thesis were true, there would be cogent inferences from nonmoral premises to moral conclusions.
4. Hence no such identity thesis is true.
5. Hence, naturalism is false.

I believe that Sturgeon has also shown this argument to be a failure, so I do not discuss it in this book.²⁵ I should note that realist-expressivism can explain why we might want to deny the third premise. There is, for example, a usage in which “redneck” is used to express contempt. On this usage, we would reject inferences from “x is a member of the white rural laboring class” to “x is a redneck,” even if “redneck” and “member of the white rural laboring class” ascribe one and the same property.²⁶ Similarly, if a moral predicate “M” is standardly used both to ascribe a

24 Sturgeon rejects the second premise, arguing that a true identity thesis need not be a conceptual truth. He also argues that the first premise is trivial: a naturalist thinks that M_{ness} is a natural property and thinks, of course, that M_{ness} is identical to M_{ness}. See Sturgeon 2006, 95–98. Smith rejects the third premise, denying that a conceptual truth must be obviously true. See Smith 2003, 194–202. See also Copp 1995a, 230–231.

25 Sturgeon 2006, 102–110.

26 Some people seem to apply the term only to white people living in rural areas of the southern United States. I quote from the *American Heritage Dictionary of the English Language*, 3rd ed., s.v. “redneck.”

property and to express a moral attitude, we might reject inferences from “x is N” to “x is M,” even if “M” and “N” ascribe the same property.

Moral naturalists need to explicate their view and to answer many significant objections and worries. In my view, the key question is whether moral naturalism can meet the normativity constraint. Jonathan Dancy argues very directly that it cannot. No natural fact is normative, he thinks, because no such fact is directly and immediately relevant to a decision about what to do in the way that a normative fact would be. Moreover, if no natural fact is normative, then no natural fact is a moral fact.²⁷ My answer to this argument is found in chapter 8.

4. SOCIETY-CENTERED MORAL THEORY

In this book I use the society-centered theory as an example of moral naturalism. I will here briefly explain the theory and deal with some questions about it that I will not be able to address in the chapters to come.

The basic idea is, I believe, intuitively appealing. We need to live in societies, and societies need to be governed by shared norms or standards in order to facilitate beneficial cooperation and coordination among their members. In light of this, it is plausible that morality has the function of making society possible by providing rules governing our lives that, when they have currency in society, enable society to meet its needs. This is the intuition that underlies the society-centered theory. A further point is that different moral codes would differ in how well their currency in society would serve this function. Given the intuition about the function of morality, it is plausible to think that a basic moral proposition, such as the proposition that it is wrong to lie, would be true only if the moral code that would best serve this function included or entailed a relevantly corresponding norm, such as a prohibition on lying.

This basic idea might, however, seem to be in tension with moral realism and with the idea that morality has normative importance. For it might suggest that morality is simply a useful set of social rules, and this might suggest in turn that our moral beliefs are simply useful fictions. And the basic idea might also suggest that morality has no ‘authority’ over our decisions and actions, for it might suggest that morality is something that rational people could view as alien, or as imposed on them.

27 Dancy 2006.

I therefore want to make clear that the society-centered theory is a version of moral realism, and I also want to show how it explains the normativity of morality. Before doing so, I need to explain the theory.

The theory has two central elements: the ‘standard-based’ account of the truth conditions of normative propositions and the society-centered account of the ‘truth-grounding’ status of moral ‘standards.’

(I) *The Standard-Based Account*

The standard-based account rests on a distinction between *moral propositions* – such as the proposition that lying is wrong – and *moral standards* – such as the standard prohibiting lying that would be expressed by the imperative, “Do not lie!” Most of us *subscribe* to the standard that prohibits lying in that, among other things, we are inhibited from lying and would feel guilty to lie, but it makes no sense to suppose that someone believes a standard (given what I mean by a standard). Standards (I stipulate) are the semantic contents expressed by imperatives, and so they are not believed, nor do they represent the world as being one way or another.

The standard-based account proposes a schema that can be used to explicate the truth conditions of moral propositions in terms of the status of relevantly corresponding moral standards. The schema is intended to be applicable to laying out the truth conditions for any kind of normative proposition. The schema says that a (pure and basic) normative proposition of type K is true if and only if a corresponding standard of type K has the K-relevant ‘truth-grounding’ status.²⁸

As applied to propositions of etiquette, for instance, the account says that a (pure and basic) proposition of etiquette, such as the proposition that it is impolite to interrupt a conversation, is true if and only if a corresponding standard of etiquette has the etiquette-relevant status. Obviously, there is room for disagreement about what this status might be. In chapter 10, I briefly discuss a proposal. The important point for present purposes is that if any basic propositions of etiquette are true, there must be some such status, for there must be some property that the standard that prohibits interrupting conversations has that distinguishes it from the

28 A ‘pure’ normative proposition of type K has no non-K-normative entailments or presuppositions (other than those given by the standard-based theory itself). A ‘basic’ normative proposition of type K ascribes a K-property to something. The proposition that Smith was wrong to steal Jones’s car is impure but basic. The propositions that cursing is rude and that theft is wrong are both pure and basic.

enormous variety of arbitrary and silly standards that have no standing in etiquette, such as the standard that prohibits calling anyone by name. It is not impolite to call someone by name although it is impolite to interrupt a conversation.

There are different kinds of normative propositions, including moral propositions, propositions of law, and certain epistemological propositions. The standard-based account leaves us with a key question regarding each of these kinds: What standing must the corresponding standards have as a condition of the truth of propositions of that kind? This question is a substantive theoretical one. It is not a question that can be settled by conceptual analysis alone.

In the moral case, the key question is this: What status must a moral standard have, in order for a moral proposition to be true? The standard-based schema says that a (pure and basic) moral proposition, such as the proposition that slavery is wrong, is true if and only if a corresponding standard – presumably the standard that prohibits slavery – has the morally relevant truth-grounding status.²⁹ The idea is that there is *some* truth-grounding status, such that when a standard enjoys that status, the fact that it does underwrites the truth of corresponding moral propositions.³⁰ There must be some property that distinguishes the standard that prohibits slavery from the multitude of standards that have no moral standing, such as the standard that prohibits wearing hats indoors. Slavery is morally wrong, but the indoor wearing of hats is not.

The standard-based account is intended to explain the normativity of normative propositions. It says that a proposition is normative only if it entails (nontrivially) the existence of a standard that has an appropriate status. For example, the proposition that lying is morally wrong entails that the standard, Do not lie, has some relevant status. Standards call for certain things to be done or to be chosen. According to the standard-based account, then, a basic normative proposition is essentially relevant to action or choice because it entails (nontrivially) the existence of a

29 In Copp 1995a (25–26), I propose a Millian account of wrongness according to which an action has the property of being wrong just in case it has the complex property of (1) being forbidden by a standard included in or entailed by the moral code with the relevant truth-grounding status and (2) a standard included in or entailed by that moral code calls for the agent of the action to feel regret for having done it, other things being equal. In this book, to simplify, I ignore clause (2) of the account.

30 In Copp 1995a, I spoke of this status as that of being *justified*, but this detail is unimportant. I am ignoring my distinction between “type-one” and “type-two” normative propositions (*ibid.*, 22–24).

standard with an appropriate status that calls for action or choice. This is what accounts for the normativity of these propositions.

(II) The Society-Centered Account of Truth-Grounding Status

Suppose we had an account of the relevant truth-grounding status of moral standards. Call this the status of being morally 'authoritative.' Given the standard-based account, we could then explain both the normative content of moral propositions and the nature of moral properties. The standard-based account says that if slavery is wrong, it is prohibited by a morally authoritative standard. This suggests that the proposition that slavery is wrong is a proposition about slavery's (at least) being prohibited by a morally authoritative standard. It also suggests that the property of being wrong is the property of (at least) being prohibited by a morally authoritative standard. To fill out the standard-based account, then, we need to develop a theory of this status, the status of being morally authoritative.

Now I think it is common sense that a society needs a 'social moral code'; that is, a society needs it to be the case that there is, among its members, a system of moral standards that is generally subscribed to and that is socially enforced and culturally transmitted. A society needs to reduce the harmfulness of conflict among its members and to give its members the security they need in order to cooperate successfully. And the currency of a moral code is, I believe, the most efficient way to achieve these ends. Other things equal, a society with a social moral code would experience less conflict, less harmful conflict among its members, and more cooperation among them, and its members would be more successful at meeting their own needs and pursuing their values than would be the case if the society did not have a social moral code. With a social moral code, a society does better than it otherwise could at meeting its need to have cooperative interaction and to avoid harmful conflict among its members.

Furthermore, I believe, it is common sense that some moral codes would serve the basic needs of society better than others. The better codes are the ones whose currency would better enable a society to meet its basic needs. Given differences in the circumstances of societies, moreover, the code that would best serve the needs of one society might be different from the code that would best serve the needs of another society. A society facing an epidemic might need the currency of a prohibition of certain kinds of behavior where, in the absence of the epidemic, it would have had no need of the prohibition. Intuitively, the prohibition would have a

kind of authoritative status in the face of the epidemic that it would lack in the absence of the epidemic.

Society-centered theory is a development of these ideas. The fact that the currency of some given moral code would better serve the basic needs of a society than would the currency of some other code means that the society has reason to prefer the first code. And I want to claim that the code that would best serve the basic needs of the society is the code that is morally authoritative. It has the relevant truth-grounding status. Standards that are included in or implied by the code underwrite the truth of corresponding basic moral propositions. That is, according to the society-centered theory, a basic moral proposition is true only if a corresponding moral standard is included in or implied by the moral code the currency of which in the relevant society would enable the society better to serve its basic needs than would the currency of other sets of rules and better than would be the case if no set of rules had currency in the society.³¹

The basic theory needs to be amended in various ways. In chapter 7, I briefly discuss the possibility that distinct moral codes could be such that their currency in a society would serve the society's needs equally well. To allow for this possibility, the theory should be amended to identify wrongness in relation to society S with, roughly, the property of being prohibited by all of the codes that are tied as 'best' for society S.³² I ignore this amendment in what follows.

What are the basic needs of a society? First, a society needs to ensure its physical integrity, to ensure that its population continues to exist. Second, it needs cooperative integrity, to ensure that there continues to be a system of cooperation among its members. This requires that it ensure internal social harmony. Third, a society needs peaceful and cooperative relationships with neighboring societies.

To meet these needs, a society must, among other things, ensure that at least the bulk of its members are able to meet their basic needs. In most circumstances, moreover, it can best meet its needs by ensuring that *all* members are able to meet their basic needs with rough equality.

31 The preceding three paragraphs follow my reasoning in Copp 1996, 252–253.

32 A more fully developed account of wrongness would be Millian in nature. See note 29, above, and see Copp 1995a, 25–26. For discussion of the amendment to deal with the possibility that several codes are tied as 'best' for a society, see Copp 1995a, 198–99; Copp 1996, 257–58. Below, I mention an amendment to deal with cases in which societies overlap and one to deal with the 'state of nature.' In Copp 1995a, I also introduce an amendment to deal with coordination problems (199–200).

There may be extreme circumstances in which a society cannot meet everyone's basic needs, or cannot equally meet everyone's basic needs. And a society may need to deter some individuals from antisocial behavior by threatening punishments through a legal system. But leaving aside issues of punishment and deterrence, and leaving aside extreme circumstances, a society best meets its needs by ensuring that its members are able to meet their needs with rough equality. For in the absence of a special reason connected to its needs, there is no reason from the standpoint of the society to favor one group over another. And people are more likely than otherwise to contribute willingly and well to the overall flourishing of the society when they have been able to meet their needs.³³

The issue of how best to promote societal needs clearly is empirical. Hence, the moral implications of the theory are both contingent and somewhat speculative. However, I think it is likely that the theory yields a deontological moral code of a familiar kind.³⁴ Since societies have basically the same needs, I think the moral codes that are authoritative relative to different societies will tend to be similar in content. Yet societies can be in different circumstances, which means that the moral codes the currency of which best meets their needs are unlikely to be exactly the same.

This account raises a number of questions and worries. What is a society? Which is the 'relevant' society in a given context? Is the theory tenable given its apparent relativistic implications? I will return to these issues.

(III) *Constructivist and Nonconstructivist Versions of the Theory*

Before addressing these issues, however, I need to distinguish two versions of the society-centered theory, the version of the theory that I have sketched so far, which is 'nonconstructivist,' and the 'constructivist' version that I originally presented in *Morality, Normativity, and Society*.³⁵ I call the latter the "original" theory and I call its nonconstructivist twin the "basic" theory. In the chapters of this book and in other recent essays, I have appealed to the basic theory rather than the original theory.

Both the basic theory and the original theory provide an account of the truth-grounding status of moral standards. The original theory

33 In the preceding two paragraphs, I follow Copp 1996, 255–256.

34 See Copp 1995a, 201–209.

35 Copp 1995a.

defines the truth conditions of moral propositions in terms of the status that corresponding standards have when the society would be *rational to choose* them to serve in the society as the societal moral code. It claims roughly that a basic moral proposition is true only if a corresponding moral standard is included in or implied by the moral code that the relevant society would be rational to choose to serve in it as the societal moral code. Now I understand the rationality of a society's choice to depend on whether the choice would best serve the society's basic needs.³⁶ Hence the two theories share the central idea that the currency of a societal moral code can serve a society's basic needs and that the currency of some codes would better serve a society's needs than the currency of some other codes. Unlike the original theory, however, the *basic* theory focuses directly on this point about societies' needs and does not rely on any claims about rational choice. It defines the truth conditions of moral propositions directly in terms of the status that corresponding standards have when their serving in a society as the societal moral code would enable the society better to serve its needs than would the currency of other sets of standards.

The original society-centered theory is 'constructivist' in that it is committed to the thesis that moral propositions are 'made true' by a relevant kind of 'endorsement' from a preferred 'standpoint.'³⁷ We can say that a constructivist theory defines an 'endorsement function' that takes a specified kind of input and yields moral propositions (or moral propositions of a certain kind) as output; a constructivist theory holds that there is an endorsement function *F* such that a moral proposition (of kind *K*) is true just in case (and because) *F* yields the proposition as 'output' given a relevant 'input.' If we take this to be the central doctrine of constructivism, the original theory is constructivist.³⁸

36 Strictly speaking, it depends on whether the choice would best serve the society's needs and enable it to serve its values. I ignore the complication about values since only a society's nonmoral values would be relevant and societies lack any interesting nonmoral values. See Copp 1995a, 190–198 and 206–207. I develop my account of rational choice, and explain the idea of a basic need, in Copp 1995a, ch. 9. I discuss the basic needs of societies at pp. 192–194.

37 Shafer-Landau uses the idea of endorsement from a preferred standpoint to explicate constructivism (2003, 14–16). He classifies the society-centered theory as constructivist (39).

38 I want to accommodate a variety of notions of what 'endorsement' comes to as well as various notions about the 'preferred standpoint.' There are naturalistic as well as nonnaturalistic forms of constructivism, so I want to explicate the idea in a way that is neutral between naturalism and nonnaturalism. There are, for example, the Rawlsian idea of choice in

It is worth mentioning that a constructivist theory can be a kind of naturalistic moral realism. It can accept the five doctrines of moral realism as well as the central doctrine of moral naturalism.³⁹ The original society-centered theory was a version of naturalistic moral realism.⁴⁰

In earlier work, I introduced the idea of a ‘practical theory.’⁴¹ A practical theory is a constructivist theory that invokes in a substantive way a conception of practical reason or of rationality or reasonableness in defining its endorsement function.⁴² Some practical theories adopt a broadly Hobbesian conception of practical reason, while others take a broadly Kantian approach to the idea of rationality.⁴³

The original version of the society-centered view is, then, a constructivist practical theory but the basic theory is not. According to the basic theory, the truth of a moral proposition depends on which system of standards is such that its currency would best serve the needs of the relevant society, which is an endorsement-independent matter.

The basic society-centered theory is simpler than the original theory. It does not presuppose a theory of practical rationality.⁴⁴ It does not depend on the idea that societies are capable of rational choice. And it is not constructivist. The basic theory entails the original theory if we assume an appropriate view about rational choice as well as the view that societies are capable in principle of choice. But since these views are controversial, it is preferable to dispense with them and focus attention on the basic theory. Since the original theory and the basic theory amount basically

the original position (Rawls 1971), Scanlon’s idea that certain moral principles could not reasonably be rejected (1998), and the idea behind at least some ‘response-dependent’ theories (D’Arms and Jacobson 2006). Divine command views are also constructivist (Quinn 2006).

39 Some philosophers take moral realism to exclude constructivism, but I think their disagreement with me is terminological. If they use “constructivism” for a view that denies one or more of the five doctrines characteristic of moral realism, then their use of “constructivism” differs from mine, and I agree that their so-called constructivism is not realist on my usage. If they use “realism” for a kind of view that adds the denial of constructivism to the five doctrines, then their use of “realism” differs from mine. Shafer-Landau adopts a usage of the latter kind (2003, 15, 17–18). I discuss Shafer-Landau’s position in Copp 2005.

40 Copp 1995a, 223–230.

41 Copp 1991.

42 In *ibid.*, I distinguished between ‘epistemic theories’ and ‘practical theories.’ I now think that the notion of an epistemic theory was unhelpful, but it is useful to have a distinction between practical theories and other kinds of constructivist theories. For a constructivist theory that is not a kind of practical theory, see the response-dependence view defended in Brower 1993.

43 Cf. Gauthier 1986 with Korsgaard 1996.

44 This was pointed out by Richmond Campbell (1997).

to different formulations of the same underlying view, since they are (nearly) extensionally equivalent,⁴⁵ since they are supported by the same arguments, and since the basic theory is simpler, the basic theory seems clearly to be preferable.

The basic theory has the further advantage of directness. Like other practical theories, the original theory aims to ‘reduce’ the normativity of morality to the normativity of rational choice.⁴⁶ It postpones the difficult question of whether it is possible to explain normativity in a fundamental way or whether, instead, normativity must be left as an unexplained primitive. The availability of the basic theory shows, however, that there is no need to rest society-centered theory on a theory of rational choice. The basic theory aims to explain the normativity of morality without a detour through the theory of rational choice.⁴⁷

(IV) Morality and Society

The concept of a society that figures in the society-centered theory plays a central role in political philosophy. John Rawls holds, for example, that the primary subject of justice is the “basic structure” of society.⁴⁸ Egalitarians are typically concerned to achieve equality among the members of society along some favored parameter.⁴⁹ Societies of course are typically organized into states, but this is not necessary, for there is, I think, a North American society even though there is not a single North American state. The example shows that societies are not necessarily ethnically or culturally homogeneous or united and that the members of a society might not identify with the group or with its history. Hence the entities I call societies must be distinguished both from states and from nations.

Societies are relatively comprehensive of the various functions and roles required for a group to be self-sufficient, and they are relatively self-sufficient. They are multi-generational both in the sense that their membership includes members of several generations and in the sense that their existence extends through several generations in time. They are territorial, and their membership includes virtually everyone residing

45 Jon Tresan pointed out that there may be exotic cases that distinguish them extensionally.

46 Copp 1995a, 54–56.

47 At one point, I believed that an adequate theory would have to be a practical theory. But I now think that this was a mistake. See Copp 1990, 1991.

48 Rawls 1971, 7–8.

49 Scanlon writes, for example: “virtually every society is marked by forms of inequality the elimination of which is a political objective of the first importance” (1997, 1).

permanently in their territories. But societies have permeable borders and their borders might not be precisely defined in the way the borders of states are defined. Membership is inherited at birth, although a person can leave one society and join another as a result of moving permanently from the territory of one society into that of another society. The people that the members of a society interact with, in securing the material necessities of their lives as well as in pursuing cultural priorities, are, by and large, also members of the society. Such interactions are governed by norms that are widely shared in the society. A society provides its members with a framework for their lives, for most of their friendships and important relationships are with other members of the society.

Most of the characteristics I have mentioned are matters of degree, and an entity might exhibit one of them to a rather large degree but another to a rather lesser degree. This needs more attention than I can pay to it here. Yet the following formula can serve as a rough definition.

A society is a multi-generational temporally extended population of persons, embracing a relatively closed network of relationships of friendship, affection, kinship, and cooperation in reproduction, and limited by the widest boundary of a distinctive and salient system of instrumental interaction that facilitates pursuit of the necessities of life and the priorities of the group's culture.⁵⁰

Societies can be nested in the way that French society is contained within European society, and they can overlap in the way that, arguably, Basque society overlaps with both French society and Spanish society. This fact requires an amendment to the underlying society-centered theory, although I ignore the amendment here.⁵¹ This fact also raises an obvious question: When we are trying to decide what we are morally obligated to do, which society is relevant? To answer this question, we need to discuss the relativistic nature of the society-centered theory.

(V) The Relativism in Society-Centered Theory

On the society-centered view, wrongness is (roughly) a relation that holds between an action and a society just in case the action is prohibited by the moral code the currency of which in the society would best serve the society's basic needs. Similarly, virtuousness is (roughly) a relation that

50 Copp 1995a, 128, and, for an extended discussion, ch. 7. In the preceding three paragraphs I draw on Copp 1997, 190–192.

51 Copp 1995a, 209–213; Copp 1996, 256–257.

holds between a trait of character and a society just in case the moral code the currency of which in the society would best serve the society's basic needs calls on people to acquire the trait, to see to it that their children have the trait, and so on. Other moral properties are similarly relational. Of course this is not obvious. It is not obvious because, in most circumstances, *context* determines which society is relevant so that we do not need to refer *overtly* to the society. Because of this, a person who is competent in the ordinary way with moral concepts might not realize that a societal parameter needs somehow to be supplied or assumed in order for a remark to the effect that, say, capital punishment is wrong, successfully to express a proposition.

There are many other cases in which people of ordinary conceptual competence might not realize that a certain predicate expression actually expresses a complex relation with a hidden parameter. For a simple example, consider the remark that Warshawski is large. She is large for a house cat, but small for a feline. Largeness is actually a relation between objects and comparison classes, where the comparison class in a given case is typically determined by the context rather than being explicitly mentioned.⁵² Because of this, one might not have realized that largeness is relational. For a more interesting example, consider the phrase, "has a temperature of 20 degrees Celsius." Given how the Celsius scale is defined, this phrase expresses a highly complex relational property; it relates something, the temperature of which is at issue, to the freezing and boiling points of water at standard air pressure, where standard air pressure is defined as the mean sea level pressure at the latitude of Paris.⁵³ It is obvious that most people who are competent in determining the air temperature by looking at a thermometer do not know what this relational property is. They might think that temperature is a rather simple property. Given these examples, we should not be surprised to find that moral predicates express properties that are relational in ways we had not anticipated.

I say, then, that moral properties are relational and that one relatum is a relevant society. The society that is relevant in a given case is typically determined by the context. The *default* is, I think, that the society

52 See Harman 1975, 3.

53 The contemporary definition of the Celsius scale is actually different from and more complex than this. I rely here on the on-line encyclopedia, Wikipedia: <http://en.wikipedia.org/wiki/Celsius>. There can be different views about the exact nature of the relation at issue depending on exactly how the semantics of the definition is understood. This point supports my claim that one can be competent with Celsius temperatures without knowing exactly what properties or relations are at issue.

that is semantically relevant when a moral claim is made or entertained is the smallest society that embraces the person making or entertaining the claim as well as all the people who are party to the conversation and all the people referred to or quantified over or whom the speakers or thinkers have in mind in the context.⁵⁴ Suppose Bush says that capital punishment is permissible. If he says this in a news conference in New York, where the intended audience is newspaper readers in the United States, then he has expressed the proposition that capital punishment is permissible in American society. But if he says the same thing in Paris at a news conference with Chirac, and if Chirac says in the same context that capital punishment is not permissible, then the two men have disagreed about the permissibility of capital punishment. The smallest society that embraces both men is perhaps the very large and loosely organized society that encompasses Europe and former European colonies, so they might best be construed as disagreeing about the permissibility of capital punishment in that larger society. If a philosopher says that capital punishment is impermissible, the relevant society is likely to be the (notional) society of all rational persons, since it is likely that the philosopher views morality as being of universal scope, prescribing duties incumbent on all rational persons. In some contexts, the default does not obtain. For example, if Bush says at the Paris news conference that capital punishment is permissible in America, his remark can be taken at face value as relativizing the claimed permissibility to American society.⁵⁵

In some cases, the smallest society that embraces the group of people at issue in a given context might be the global society of all persons. In other cases, however, the group of people at issue might not be contained within any society. To deal with such ‘state of nature’ cases, I propose a minor amendment to the theory. I say that the duties of people in a group that is not part of any society are determined by the content of the moral code the currency of which in the group would best contribute to its acquiring the properties of a society – including its coming to be characterized by a system of cooperation – and then to serving its basic needs.⁵⁶

According to the account I am giving here, when we wonder what we morally ought to do in a given situation, our intentions and the context will determine which society is the relevant one. Of course, whether we

54 Copp 1995a, 221.

55 In the preceding three paragraphs, I follow Copp 1997.

56 Ibid., 198–199. I made a somewhat different suggestion in Copp 1995a, 121–122.

reach a correct conclusion about what ought to be done will depend in part on how closely our actual moral convictions approximate the convictions we would have if we subscribed to the moral code the currency of which in the relevant society would best enable it to meet its basic needs. This raises issues in moral epistemology, which I discuss in chapters 2 and 3.

The relativism in the society-centered view is actually quite weak. There are various details that mitigate the variability of morality from society to society. For example, there is, I believe, a global society, and in many contexts the global society is the relevant society. Moreover, since societies have the same needs, there is reason to expect that the moral code that is authoritative relative to one society will be similar to the moral code that is authoritative relative to another society. If capital punishment is wrong relative to French society but permissible relative to American society, then the societies must be in somewhat different circumstances.

(VI) *Objections*

A variety of objections could be raised to the society-centered theory, and I respond to many of them in this book. As I mentioned before, the basic idea of the theory might seem to be at odds both with moral realism and with the normativity of moral judgment. I hope I have already said enough to make it clear why I think these worries are ill-founded. In chapter 5, I respond in more detail to the worry about realism, and I respond to the worry about normativity in chapters 8 through 10.

One might worry that the society-centered view conflicts with our moral intuitions. I do not say very much about this issue in this book.⁵⁷ I do, however, discuss the idea that the society-centered view cannot account for our intuition that certain kinds of action are wrong *period*. One might think, for example, that slavery is wrong under all circumstances, and in particular that slavery would have been wrong even if our society had been in circumstances where its needs would have been served better by the currency of a moral code that permitted slavery. Of course this objection could be interpreted simply as a denial of the society-centered theory rather than as the beginning of an argument against the theory. But it can also be interpreted as expressing a moral intuition that is in apparent

57 Anton Leist pressed the worry (1997) and I have responded to it (Copp 1995a, 201–209, 213–216; 1997, 203–212; 1998, 128–139).

conflict with the implications of the theory. In chapter 4, I argue that the society-centered theory can in fact account for the intuition that slavery is wrong period. I think that the theory does much better at accounting for our moral intuitions than one might initially believe.

5. AN OVERVIEW OF THE BOOK

The book is divided into three parts. Part **One** contains four chapters on the nature of moral naturalism and on metaphysical and epistemological issues. Part **Two** contains three chapters on semantic challenges that face moral naturalism. Part **Three** contains three chapters on the normativity of moral judgment and on the relation between morality and rationality.

Chapter 1, “Why Naturalism?” introduces an epistemological characterization of the idea of a natural property and offers an argument that moral properties are natural properties on this characterization. On this characterization, a naturalist should hold that all substantive moral knowledge is empirical. But certain common sense observations can appear to undermine the idea that our moral knowledge is empirical. For instance, sound moral belief seems to be a result of thinking sensitively about an issue rather than a result of empirical theorizing. In chapter 2, “Four Epistemological Challenges to Ethical Naturalism,” I discuss challenges of this kind. I argue that the observations that fuel the challenges are actually compatible with moral naturalism. I also address the worry that the society-centered view would seem to imply that moral knowledge is a kind of sociological knowledge, which is not true to our ways of coming to decisions about moral questions. In chapter 3, “Moral Naturalism and Self-Evident Moral Truths,” I show that moral naturalism is compatible with the idea that many moral propositions are self-evident.

The final chapter in Part **One**, “Moral Necessities in a Contingent World,” addresses the intuitively plausible idea that certain substantive moral propositions are necessary truths in a relevant sense of the term. We might have the intuition, for instance, that slavery is wrong period. One might think that this idea cannot be accommodated by moral naturalism. In this chapter, I argue on the contrary that only a naturalistic theory that provides a substantive and rich account of the nature of moral properties can explain what grounds the moral necessities.

The three chapters in Part **Two** deal with central issues in the defense of moral naturalism against noncognitivism. In chapter 5, “Realist-Expressivism,” I address the issue of whether moral predicates ascribe

properties or whether, instead, they are used to express conative attitudes such as approval or disapproval. I point out that moral predicates may be used both to ascribe properties and to express attitudes. This proposal can help to explain some intuitions that lead people to believe that judgment internalism is true. It can also explain the semantic difference between a moral predicate and the naturalistic predicate, if there is one, that could be used to ascribe the very property that is also ascribed by that moral predicate.

Any version of moral naturalism must explain how moral predicates come to refer to natural properties. To explain this, a naturalist needs to adopt a general semantic theory about the relation between predicates and properties and then apply the theory to moral predicates. Terence Horgan and Mark Timmons have proposed an argument – the Moral Twin Earth argument – that is intended to show that this strategy cannot work.⁵⁸ The argument is intended, more generally, to undermine the idea that moral predicates are used to ascribe properties. My reply to this argument is in chapter 6. Horgan and Timmons have responded that the only way to avoid their objection is to leave it indeterminate which natural property is ascribed by a given moral predicate.⁵⁹ In chapter 7, I aim to show that they are mistaken by sketching a semantic theory that meshes smoothly with the society-centered theory and that determines which natural properties are ascribed by given moral predicates.

Part [Three](#) of the book contains three chapters that deal in one way or another with the normativity of moral judgment. Chapter 8, “Moral Naturalism and Three Grades of Normativity,” is my most thoroughgoing attempt to address the normativity constraint. Since philosophers seem to have different things in mind when they speak of morality as “normative,” I distinguish three ‘grades of normativity.’ For each of these grades, I investigate the plausibility of the idea that morality has that grade and the viability of naturalistic accounts of it. It is sometimes thought, for example, that an adequate account of the normativity of morality must show that it gives rise to reasons that any rational person would take into account in deliberation, if he or she were aware of them, just in virtue of being rational. I argue that this is a mistake.

It is sometimes thought that an adequate account of morality must show that moral considerations ‘override’ normative considerations of

58 See, e.g., Horgan and Timmons 1992.

59 Horgan and Timmons 2000.

other kinds, such as reasons of self-interest. In chapter 9, “The Ring of Gyges,” I argue that moral reasons are not overriding. Morality is not a supreme normative standpoint that directs and orders the realm of reasons. Indeed, there is no such standpoint. Hence, it is a mistake to think that moral considerations have ‘final authority’ over our decisions and actions – unless we have in mind the uninteresting idea that they have final moral authority. I argue that moral reasons do not override reasons of self-interest, but neither do reasons of self-interest override moral reasons. There simply are different kinds of reasons.

This position gives rise to awkward questions about the relations among reasons of the different kinds. Moreover, since we need to make decisions, we want to know what we have most reason to do. In addition to these practical issues, there are also metatheoretical issues regarding our thought and discourse about nonmoral reasons that are analogous to the metaethical issues about moral reasons that the society-centered theory is designed to answer.

I take up these issues in chapter 10, “The Normativity of Self-Grounded Reason.” I argue in this chapter that ‘self-grounded reasons’ – roughly, reasons to serve one’s values – are such that any rational person would then take into account in deliberation, if he or she were aware of them, just in virtue of being rational. Moral reasons are not *necessarily* self-grounded, but they are self-grounded in morally good people, for good people have moral values and rationally guide their lives in accord with their values. I argue, moreover, that given the nature of practical deliberation, if a rational agent with relevant knowledge and understanding reaches a decision about what to do as a result of deliberation, then, other things being equal, his or her decision is to act in accord with his or her values. That is, self-grounded reasons have ‘priority in deliberation.’ I argue in chapter 9 that self-grounded reasons do not have a metaphysical status in virtue of which they override reasons of other kinds. If I am correct, however, they do have ‘priority in deliberation.’

It is widely thought that moral naturalism is hopeless, and that if we are to avoid nihilism or the error theory, we must take up either nonnaturalism or noncognitivism. My goal in this book is to answer many of the standard objections to moral naturalism, including especially the objection that moral naturalism cannot explain the normativity of moral thought and discourse. In the course of defending moral naturalism, I also aim to show that the society-centered theory is a plausible version of moral naturalism. In the end, after the objections and challenges and responses and counterresponses have all been registered, there will of course be no

decisive argument either for moral naturalism or against it. Yet at this point in the ongoing argument, I believe that moral naturalism remains in a strong position.

REFERENCES

- Blackburn, Simon. 2006. "Antirealist Expressivism and Quasi-Realism." In Copp 2006b: 146–162.
- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, N.Y.: Cornell University Press.
- Brink, David O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Brower, Bruce. 1993. "Dispositional Ethical Realism." *Ethics* 103: 221–249.
- Campbell, Richmond. 1997. "Critical Notice of David Copp, *Morality, Normativity, and Society*." *Canadian Journal of Philosophy* 27: 423–444.
- Copp, David. 1990. "Explanation and Justification in Ethics." *Ethics* 100: 237–258.
- . 1991. "Moral Skepticism." *Philosophical Studies* 62: 203–233.
- . 1995a. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1995b. "Moral Obligation and Moral Motivation." In *On the Relevance of Metaethics: New Essays on Metaethics*, ed. Kai Nielsen and Jocelyne Couture. *Canadian Journal of Philosophy*, supp. vol. 21: 187–219.
- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In *Moral Knowledge? New Readings in Moral Epistemology*, ed. Walter Sinnott-Armstrong and Mark Timmons, 243–266. New York: Oxford University Press.
- . 1997. "Does Moral Theory Need the Concept of Society?" *Analyse und Kritik* 19: 189–212.
- . 1998. "Morality and Society – The True and the Nasty: Reply to Leist." *Analyse und Kritik* 20: 30–45.
- . 2005. "A Skeptical Challenge to Moral Non-Naturalism and a Defense of Constructivist Naturalism." *Philosophical Studies* 126: 269–283.
- . 2006a. "Introduction: Metaethics and Normative Ethics." In Copp 2006b: 3–35.
- , ed. 2006b. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
- Dancy, Jonathan. 2006. "Nonnaturalism." In Copp 2006b: 122–145.
- D'Arms, Justin, and Daniel Jacobson. 2006. "Sensibility Theory and Projectivism." In Copp 2006b: 186–218.
- Darwall, Stephen. 1983. *Impartial Reason*. Ithaca, N.Y.: Cornell University Press.
- Darwall, Stephen, Allan Gibbard, and Peter Railton. 1992. "Toward *Fin de siècle* Ethics: Some Trends." *Philosophical Review* 101: 115–189.
- Frege, Gottlob. 1984a. *Collected Papers on Mathematics, Logic, and Philosophy*, ed. Brian McGuinness, trans. M. Black, V. Dudman, P. Geach, H. Kaal, E.-H. W. Kluge, B. McGuinness, and R. H. Stoothoff. Oxford: Basil Blackwell.
- . 1984b. "On Concept and Object." In Frege 1984a: 182–194.
- . 1984c. "On Sense and Meaning." In Frege 1984a: 157–177.
- . 1984d. "Thoughts." In Frege 1984a: 351–372.

- Gauthier, David. 1986. *Morals by Agreement*. Oxford: Oxford University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Harman, Gilbert. 1975. "Moral Relativism Defended." *Philosophical Review* 84: 3–22.
- Horgan, Terence, and Mark Timmons. 1992. "Troubles on Moral Twin Earth: Moral Queerness Revived." *Synthese* 92: 221–260.
- . 2000. "Copping Out on Moral Twin-Earth." *Synthese* 124: 139–152.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge: Cambridge University Press.
- Leist, Anton. 1997. "For Society – Against Morality? On David Copp's Attempt to Put Society at the Centre of Ethics." *Analyse und Kritik* 19: 213–228.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Miller, Alexander. 2003. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt. Cambridge: Cambridge University Press, 1993.
- Quinn, Philip L. 2006. "Theological Voluntarism." In Copp 2006b: 63–90.
- Railton, Peter. 1986. "Moral Realism." *Philosophical Review* 95: 163–207.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Scanlon, T. M. 1997. "The Diversity of Objections to Inequality." The Lindley Lecture, University of Kansas, 1996. Lawrence: Department of Philosophy, University of Kansas.
- . 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. Oxford: Clarendon Press.
- Smith, Michael. 2003. "Moral Realism." In *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-ethics*, 181–207. Cambridge: Cambridge University Press.
- Sturgeon, Nicholas L. 1984. "Moral Explanations." In *Morality, Reason, and Truth*, ed. David Copp and David Zimmerman, 49–78. Totowa, N.J.: Rowman and Allanheld.
- . 2006. "Ethical Naturalism." In Copp 2006b: 91–121.

Part One

Naturalism: Epistemology and Metaphysics

Why Naturalism?

My goal in this chapter is to explain what ethical naturalism is, to locate the pivotal issue between naturalists and nonnaturalists, and to motivate taking naturalism seriously. It is no part of my goal to establish the truth of naturalism. There are various familiar objections to it, including, most importantly, the objection that naturalism cannot explain the normativity of moral judgment. In this chapter, my goal is simply to motivate naturalism sufficiently that the attempt to deal with the objections will seem worthwhile.

An ethical naturalist holds that there are moral properties and relations – for example, there are moral rightness, goodness, justice, and virtuousness – and she holds that these properties and relations are ‘natural.’¹ Accordingly, when a naturalist hears us say that something is right or wrong, just or unjust, she takes the truth of what we say to depend on whether the relevant thing has the relevant property, and she takes this to depend in turn exclusively on the way things are in the natural world. The chief problem, of course, is to explain what it might mean to claim that moral properties are *natural* properties. I think that once this is properly explained, naturalism will seem enormously attractive. If we

I have discussed the issues in this chapter with many people, including Janice Dowell, Thomas Hofweber, Jeffrey C. King, Peter Railton, Michael Ridge, Gideon Rosen, Sydney Shoemaker, David Sobel, and Jessica Wilson. I am grateful to all of them for their help and patience. I presented drafts of the chapter to the 2001 meetings of the North American Society for Social Philosophy; to the departments of philosophy at Washington University in St. Louis, Tulane University, and the University of Michigan, Ann Arbor; to the 2002 meetings of the British Society for Ethical Theory; and to the Simon Fraser University Conference in Honour of Steven Davis. I am grateful to members of all of these audiences for helpful discussion.

1 In what follows, I ignore relations.

believe that there are moral properties at all, we will find ourselves moved in the direction of naturalism.

It is important to understand from the outset the relation between ethical naturalism and the unrestricted form of naturalism according to which the natural world is all that there is. There are two points. First, an ethical naturalist can consistently reject an *unrestricted* naturalism. She needn't be a naturalist about mathematics, for example. Her thought is specifically about morality. Second, unrestricted naturalism does not commit one to ethical naturalism. Ethical naturalism, as I understand it, involves the thesis that there are *moral properties*, and this thesis is not entailed by unrestricted naturalism. Ethical naturalism is a kind of moral *realism*, but an unrestricted naturalist could accept an error theory in ethics, or a version of expressivism.²

Moral realism is controversial, and so might be my characterization of it as involving the thesis that there are moral properties. Nevertheless, I will not attempt here to defend either moral realism or my characterization of it. Instead, I will simply take moral realism as given and proceed to investigate the naturalistic form of realism. Ethical naturalism combines moral realism with the doctrine that the moral properties are 'natural.' This is where I want to focus my attention. What does the naturalist mean by a natural property?

1. G. E. MOORE ON NATURAL PROPERTIES

Since G. E. Moore famously argued against moral naturalism, I want to begin by asking what Moore meant by "naturalistic ethics."³ In *Principia Ethica*, in what appears to be his official characterization, Moore says that a naturalistic theory selects "some one *natural* property" and proposes that "to be 'good' means to possess the property in question."⁴ For example, a hedonistic theory might propose that to be 'good' means to be pleasant. Moore's way of formulating his view does raise some questions, but he appears to have understood ethical naturalism in the way that I am proposing to understand it, as the thesis that moral properties are natural. The

2 Mackie 1977, ch. 1; Gibbard 1990.

3 Moore 1903. The naturalistic fallacy is introduced in section 10, at p. 62; the open question argument in sec. 13, at pp. 66–68. The argument and the fallacy are linked to "naturalism" in sec. 14, at pp. 71–72. We cannot equate naturalism with the kind of view Moore thought guilty of the naturalistic fallacy because he held that "metaphysical ethics" also commits the fallacy (sec. 25, p. 91).

4 Ibid., sec. 26, p. 91.

key problem, then, is to determine what he meant by a “natural property.” Moore made several different suggestions about this, but years later, in the preface to the second edition of *Principia*, which he never published, and also in his “Reply to My Critics,” Moore acknowledged that his attempts to explain the idea of a natural property in *Principia* were “hopelessly confused.”⁵ He had *no* adequate way of distinguishing natural from nonnatural properties, and this means that he had no principled way of distinguishing the “naturalism” he denied from the “nonnaturalism” he advocated.

In my view, the most promising of Moore’s characterizations of naturalism was the first one he gave in *Principia*. He suggested there that, according to naturalistic ethics, “Ethics is an empirical or positive science: its conclusions could all be established by means of empirical observation and induction.”⁶ Now it clearly would be a mistake to hold that ethics is a science. Instead, the naturalist ought to say that ethics is *empirical* in the sense that any ethical knowledge is based in “empirical observation and induction.” It is no part of naturalism to deny that we could have nonempirical knowledge of conceptual moral truths or analytic moral truths. Hence, the naturalist should say that any knowledge we have of *synthetic* moral truths must be empirical. And, in line with this, she might propose that moral *properties* are empirical in that any knowledge we have of synthetic propositions about their instantiation must be empirical.

Moore is of course a nonnaturalist on this proposal. He says that to know which things have intrinsic value, “it is necessary to consider what things are such that, if they existed *by themselves*, in absolute isolation, we should yet judge their existence to be good.”⁷ This certainly does not seem to be an *empirical* method. Among the things we can know in this way, Moore suggests, is that friendship is good,⁸ which is surely a synthetic proposition. Moore therefore appears to think that our fundamental moral knowledge is *nonempirical* knowledge of *synthetic* moral truths. On the proposal we are considering, moreover, he views goodness as a nonnatural property since he thinks that we can have nonempirical knowledge of the synthetic truth that goodness is instantiated in friendships. On the current proposal, then, Moore qualifies as a nonnaturalist, which, of course, is the result we wanted to find. I will return to this proposal after considering some contemporary conceptions of the natural.

5 Moore 1993b, 13; Moore 1968, 582.

6 Moore 1903, sec. 25, p. 91.

7 Ibid., sec. 112, p. 236, see also sec. 55, p. 145.

8 Ibid., sec. 113, p. 237. Moore writes of the intrinsic value of “the pleasures of human intercourse.”

2. CONTEMPORARY CONCEPTIONS OF NATURAL PROPERTIES

The idea of the natural is central to philosophical debates in many areas of philosophy. Since different conceptions may be appropriate in different areas, I see no reason to tie our understanding of ethical naturalism to the best understanding of the natural in any other area. It will nevertheless be useful to consider proposals that are familiar from the literature, even if only briefly. There seem to be four basic approaches: (A) reductionist proposals, (B) ostensive definitions, (C) metaphysical definitions, and (D) epistemic definitions.

(A) *Reductionist and Relational Approaches.* A reductionist or relational strategy specifies a base class of properties that are supposed to be uncontroversially natural and then requires, of any other property that is to qualify as natural, that it be suitably related to properties in the base class. As I understand “reductionism,” a naturalistic reduction of moral properties would involve producing necessary and sufficient conditions for the correct predication of a moral predicate, with the conditions being specified in a preferred terminology, using only predicates taken to refer to properties in the base class.⁹ Other kinds of relational approach seek to relate moral properties to properties in the base class directly by means of a metaphysical relation, such as supervenience.

These proposals are problematic. For one thing, even nonnaturalists must agree that moral properties supervene on nonmoral natural properties. Moore agrees, for example.¹⁰ For another thing, some naturalists would deny that there is a need to reduce moral properties to properties in an independently specified base class.¹¹ The availability of such a view means that goodness might be a natural property even if Moore is correct that it is simple and unanalyzable.

The key problem, however, is that approaches of these kinds do not tell us what it is about a natural property that makes it natural. They depend on the prior selection of a base class of natural properties, and they are therefore parasitic on a proposal of some *other* kind for an account of what distinguishes natural from nonnatural properties. Of course, if we had such an account, we could apply it directly to the moral properties to test

⁹ This discussion follows King 1994, 55–56.

¹⁰ Moore 1993a, 292–293. Moore does not use the term “supervenience.” See also Moore 1968, 587–588. Frank Jackson argues that the supervenience of moral properties on natural properties entails that moral properties are natural (1998, 118–125).

¹¹ Sturgeon 1984; Miller 1985; Railton 1989, 161.

whether they are natural properties, just as we would test the properties in the base class.

(B) *Ostensive Definitions*. The strategy here is to take the ‘natural objects’ to be such as we find around us, and then to define natural properties in terms of the natural objects. Following Frank Jackson, we could proceed by “pointing to some exemplars” of ordinary objects, such as “tables, chairs, mountains, and the like,” and then say that natural properties are those that are “needed to give a complete account of things like them.”¹² However, even a nonnaturalist can insist that moral properties must be mentioned in or implicated by a “complete account” of ordinary objects. A naturalist would insist, of course, that she has in mind what is needed in a complete ‘naturalistic account’ of the world, but this proposal simply shifts the burden to the idea of a ‘naturalistic account.’ Some naturalists would suggest that by a naturalistic account they mean a scientific account. On approaches of this kind, however, it appears that we could simply take the natural world to be the world studied by the sciences, and there would be no need for the ostensive part of the proposal.

(C) *Metaphysical Characterizations of the Natural*. Since the category of natural properties is meant to be metaphysical, it would be ideal if we could define the natural in metaphysical terms. The literature offers at least four suggestions.

(1) Natural properties are sometimes said to be “descriptive characteristics” or “factual properties.”¹³ This kind of characterization of the natural is familiar, and indeed Moore flirted with it.¹⁴ The problem is that, in an ordinary sense of the word, we describe a person in saying, for example, that she is a good person. We might describe Mother Theresa as “compassionate,” for instance, and if we think she was compassionate, we ought to agree that she was ‘in fact’ compassionate.¹⁵ Nonnaturalists ought to agree with this, moreover, so the present characterization of the natural does not capture a view they reject.

(2) The natural world is sometimes said to be the causal order – the universe of events and states of affairs that are linked in a causal order.¹⁶ There are two problems with this suggestion. First, it cannot be assumed that the natural order is causal. It is sometimes said that there is no causation

12 Jackson 1998, 7.

13 Hare 1952, 145, 82; Goldman 1994, 301. See Moore 1968, 591.

14 Moore 1993a, 296–297; Moore 1968, 590–591; Hare 1952, 147, 154–155, 171–172.

15 For this usage, see Gibbard 1990, 9.

16 For a related suggestion, see Goldman 1994, 302.

at the most fundamental level of natural reality. Whether this is correct is an empirical issue. Second, a great many people hold supernatural or superstitious views about the causal order, and these views are not all naturalistic. For example, people who accept the story of creation in *Genesis* as the literal truth hold that God caused the world to exist. It would muddy the water to take their view as naturalistic.

The problem here is not that the *Genesis* view makes room for God, for there are theological views that plausibly are naturalistic. Pantheism, for instance, is a kind of naturalism about God. The problem is that the process by which God created the world in the *Genesis* view was not a *natural* process. Hence, if we are to explain the natural world in terms of the idea of a causal order, we need to develop an account of *natural* causation. The obvious way to proceed would be to explain natural causation in terms of science, as causation under scientific law. But on this approach, it appears that we could simply treat the natural world as the world that is studied by the sciences and there would be no need to invoke the idea of a causal order.

(3) Following David Armstrong, we might take the natural world to be the “spatiotemporal manifold,” the conjunction of all states of affairs in space and time.¹⁷ One problem with this proposal is that, on various Platonist conceptions of properties, *properties* are not in space-time, and this ought to be compatible with the thesis that there are *natural* properties. Of course, the objects that instantiate natural properties are in space-time, but so are the objects that instantiate moral properties. One way around this problem would be to say that the natural world consists of the spatiotemporal manifold along with the properties that are “needed to give a complete [naturalistic] account” of everything in the manifold. But as we have already seen, this approach would place a great deal of weight on the idea of a ‘naturalistic account’ of the world. On this approach, moreover, it appears that there would be no need to invoke the idea of the spatiotemporal manifold.

(4) The final metaphysical suggestion is that the natural world is the *material* or the *physical* world.¹⁸ Since this proposal involves explicit reference to science, it anticipates the discussion about scientific accounts of the world that we have so far left unfinished. The problem with this proposal for our purposes is that ethical naturalism need not be materialist or physicalist.

17 Armstrong 1989, 99, 76.

18 According to David Papineau, this is to be understood “not in terms of current physics,” but in terms of the science that eventually explains “the behaviour of matter” (1993, 2).

(D) *Epistemological Characterizations of the Natural*. Given what we have said, the obvious next idea is to take the natural world to be the world studied by the sciences and to take natural properties to be those that are “needed to give a complete [scientific] account” of this world. Proposals that tie the conception of the natural to science are very common throughout philosophy.¹⁹ Stephen Darwall, Allan Gibbard, and Peter Railton have said, for example, that the moral naturalist aims to effect an “assimilation” between ethics and science.²⁰

The reason we are tempted turn to science in explicating naturalism, I think, however, is that we take science to be our most reliable source of empirical knowledge. But if there are nonscientific means of acquiring empirical knowledge, whatever we learn by using these means ought also to count as knowledge of the natural world. And it does seem that we can have empirical knowledge that is not scientific, including knowledge of street names, dollar bills, aches and pains, and popular foods. Accordingly, we would need a rationale for tying our understanding of naturalism to science rather than to the empirical. If mental properties are epiphenomenal, for example, they presumably will play no role in the true scientific story of the world, yet for my purposes they ought to count as natural.²¹ Naturalism is not scientism, and here is the place to distinguish the two by insisting that the naturalist holds only that natural properties are empirical, not that they must be properties that figure in scientific theory. Given these considerations, I propose to explicate naturalism in terms of the empirical and to leave aside issues about the relation between science and other putative sources of empirical knowledge.

3. NATURAL PROPERTIES AS EMPIRICAL PROPERTIES

My proposal distinguishes natural from nonnatural properties on the basis of the nature of our epistemological access to them. It construes a natural property as an empirical property; that is, ignoring certain complexities, and as a first approximation, it holds that

A property is natural if and only if any synthetic proposition about its instantiation that can be known, could only be known empirically.²²

19 Kornblith 1994; King 1994, 53–56.

20 Darwall, Gibbard, and Railton 1992, 126, 165.

21 I owe this example to Jeffrey King (in personal discussion).

22 A proposition counts as being “about the instantiation” of a property if either (a) it implies that the property is instantiated or (b) it implies a proposition about the circumstances in which the property would be instantiated. The proposition that friendship is good illustrates

On my proposal, ethical naturalism is the position that moral properties are empirical properties.

The empirical is traditionally contrasted with the *a priori*, such that all of our knowledge and all warranted beliefs are either empirical or *a priori*. The basic idea of course is that *empirical* knowledge or warrant is grounded or based in *experience*. The proposal, then, is to define naturalism in terms of what we can know or believe with warrant on the basis of experience.²³

The restriction to ‘synthetic’ ethical propositions is required since, as I explained before, it is no part of ethical naturalism to deny that there might be conceptual or analytic truths in ethics and that we could have *a priori* knowledge of their truth. It is arguable, for example, that the concept of murder is the concept of a wrongful killing, and if so, then the proposition that murder is wrong is a conceptual truth, and we can have *a priori* knowledge that it is true. More significantly, it is arguably a conceptual truth that the moral supervenes on the nonmoral.²⁴ If so, then a naturalist can concede that we can have *a priori* knowledge that the moral supervenes on the nonmoral.

The empirical conception of naturalism seems to me to answer to the fundamental intuition behind naturalism. The primitive intuition is that the natural world is the world around us, the world we are immersed in. The key problem is to explain the sense in which we are ‘immersed’ in it. My proposal is to explain this on the basis of the nature of our epistemic access to the world.²⁵ The natural world is the world we know empirically. In moral theory, the empirical conception of naturalism seems to be assumed in recent work defending naturalistic forms of moral realism. For example, the debate about moral explanations that took place in the 1980s was animated by Gilbert Harman’s account of empirical inference as consisting in inference to the best explanation. In arguing that there are genuine moral explanations of empirical phenomena, Nicholas Sturgeon

(b). [Author’s note: In chapter 3, I stipulate that a proposition is *synthetic* just in case it is neither a conceptual truth nor the negation of a conceptual truth. In chapter 4, I discuss the idea of a conceptual truth.]

23 One might object that if there are facts about the way the world actually works that could not be believed with warrant by beings who are equipped the way humans are actually equipped, not even in principle, they still should be counted as natural facts. But if we suppose there are such facts, the question to ask is whether, if we had knowledge of them, it would be empirical. I owe this objection to Jeffrey King. It is avoided by the preferred formulation of my view, which I present in the [next section](#) of the chapter.

24 Jackson 1998, 125.

25 Cf. Kitcher 1992, 57–59.

was supporting a kind of naturalistic moral realism.²⁶ Similar conceptions of naturalism are found in the work of Richard Boyd, David Brink, Richard Miller, and Peter Railton.²⁷

4. THE SYNTHETIC A PRIORI

Given the traditional contrast between the empirical and the a priori, the naturalist's thesis that all of our knowledge of synthetic moral truths is empirical commits the naturalist to denying the synthetic a priori in ethics. The naturalist is committed to rejecting the possibility of a priori knowledge of synthetic moral truths, while the nonnaturalist is committed to embracing this possibility. The nonnaturalist is therefore allied with Moore and Kant, while the naturalists are lined up on the other side. The pivotal issue is whether there can be synthetic a priori knowledge in ethics.

Unfortunately, both the analytic/synthetic distinction and the distinction between the empirical and the a priori are contested. A half century ago, Quine famously argued against both the analytic/synthetic distinction and the idea that there can be a priori knowledge. More recently, from the side of a neo-rationalism, Laurence Bonjour has argued that no significant empirical knowledge is possible unless we have synthetic a priori knowledge of fundamental logical and epistemic principles.²⁸ Fortunately, there is no need to enter into the deepest of the controversies. We can set aside Quinean worries, for I do not want to rest my case for naturalism on Quinean considerations. Moreover, even if we agree with Bonjour's views about empirical knowledge, it does not follow that there can be synthetic a priori knowledge in *ethics*. So the central issue remains untouched.

The most interesting questions are about the distinction between the empirical and the a priori. Perhaps some will follow Quine in denying that there are any conceptual truths or analytic propositions. But then, on the usage I will employ, they hold that all propositions are synthetic, and it still remains to consider whether there can be a priori knowledge in ethics.

There are, unfortunately, different ways of drawing the distinction between the empirical and the a priori, and I see no reason to think

26 Sturgeon 1984; Harman 1977, chs. 1 and 2.

27 Boyd 1988; Brink 1989; Miller 1985; Railton 1989.

28 Quine 1961; Bonjour 1998, 1–6.

that there is only one *correct* way of drawing it, or to think that the same way of drawing it will be most useful in every issue in philosophy.²⁹ This means there are many different conceptions of natural properties, each of which construes a natural property as ‘empirical.’ I need to draw the line between the empirical and the a priori in a place that corresponds cleanly to the line of controversy between naturalism and nonnaturalism in ethics. Of course, the issue of what to *call* the line is unimportant. I will continue to speak of the empirical and the a priori, but nothing turns on this.

The traditional view is that whatever we know empirically, or are warranted empirically to believe, we believe on the basis of experience; a priori knowledge and warranted belief is said to be ‘independent’ of experience – except for the experience required to understand the proposition in question. Different ways of drawing the distinction between the empirical and the a priori interpret this traditional formulation in different ways. Most important is that there are different ways to understand the notion that empirical belief is ‘based’ in experience. The naturalist needn’t hold that all significant ethical knowledge or warranted belief is based in any *direct* way in experience. Instead, what she ought to say is that all ethical knowledge or warranted belief is ‘answerable’ to experience.

This idea can be clarified using definitions proposed by Hartry Field. He says,

Let’s define a *weakly a priori* proposition as one that can be reasonably believed without empirical evidence; an *empirically infeasible* proposition as one that admits no empirical evidence against it; and an *a priori* proposition as one that is both weakly a priori and empirically infeasible.³⁰

For clarity, I will call Field’s third kind of proposition “strongly a priori.”

Now there is a dispute in the literature as to whether any genuine a priori knowledge or warranted belief could be undermined by experience.³¹ In Field’s terms, the debate turns on which of the two notions, that of a weakly a priori proposition or that of a strongly a priori proposition, is theoretically more useful or closer to the traditional conception of the a priori.³² This issue is not important for our purposes. I will argue

29 A similar point is made in Boghossian and Peacocke 2000a, 3.

30 Field 2000, 117. Field emphasizes that by “reasonable” he means “epistemically reasonable” (117 n. 2).

31 See Kitcher 1992, 77 n. 79, and Bonjour 1998, 10–11. Boghossian and Peacocke provide a very brief overview of the dispute (2000a, 4–5).

32 Field 2000, 117, 119, 120–124.

that the notion of the strong a priori is the one we need to explicate the debate between naturalists and nonnaturalists in ethics.

Accordingly, I will propose understanding the ethical naturalist to deny that any synthetic proposition about the instantiation of a moral property is *strongly* a priori. As I will explain, a naturalist can agree that some substantive moral propositions can reasonably be believed *without empirical evidence*, so she can say that some such propositions are *weakly* a priori. However, she will hold that all substantive moral propositions are *answerable* to experience. They are empirically defeasible, and so they are not *strongly* a priori. In the [next section](#) I will explain why I say this, but before doing so, I need to tinker with my formulation of the empirical conception of naturalism and to attend to the idea of empirical defeasibility.

On the formulation I proposed before, the naturalist holds that a moral property is a natural property in the sense that synthetic propositions about its instantiation can only be known empirically. If we define an a priori proposition to be one that can be known a priori, we could say that, according to the naturalist, no synthetic proposition about the instantiation of a moral property is a priori. Armed with Field's distinctions, however, we now can distinguish between strongly and weakly a priori propositions, and this suggests a corresponding distinction between conceptions of a natural property. I will be using the idea of a natural* property, where

A property N is natural* if and only if (a) it is possible for N to be instantiated and (b) there are propositions about the instantiation of N that are both synthetic and possibly true, and, (c) no such proposition is strongly a priori.³³

Accordingly, I propose understanding the naturalist to hold that moral properties are natural* properties; that is, she denies that any synthetic propositions about their instantiation are strongly a priori.

Strongly a priori propositions are empirically indefeasible in that, as Field says, they “do not admit empirical evidence against them.” As Field points out, however, the idea of empirical indefeasibility needs to be interpreted with some caution because of issues raised by testimonial evidence.³⁴ Issues of this kind will be important in what follows. Suppose, for example, that a mathematician, Matty, discovers what she takes to be

33 [Author's note: This footnote was added in 2006.] Michael Ridge has pointed out in correspondence that I may need to modify this definition in order to deal with a class of problem cases.

34 Ibid., 118. Field refers to arguments about mathematics in Kitcher 1983.

a proof in system S of a new theorem T. In this case, the proposition that T is a theorem in S qualifies as weakly a priori since it is reasonably believed by Matty even though she has no empirical evidence that it is true. Suppose, however, that Matty is very insecure and her colleagues are skeptical of her proof. The skepticism of her colleagues eventually leads her to think that the proof is unsound. In this case, it seems, she would no longer be reasonable to believe that T is a theorem. And if the skepticism of her colleagues counts as empirical evidence against the proposition that T is a theorem, then the proposition is not strongly a priori. Something has gone wrong, however, for propositions of this kind surely ought to count as empirically indefeasible if any do.

We cannot avoid the problem simply by ruling out testimonial evidence. The testimony of others clearly can count as empirical evidence against beliefs about many things. I take it that epistemic norms are justified in light of the fact that conformity to them contributes to achieving the epistemic goals of gaining truth and avoiding falsehood in belief. These goals would be well served by a norm requiring that we assign less credibility than we otherwise would to propositions we believe, with which we see that others disagree, in cases in which we have no independent reason to think we are in a better epistemic position than they are.³⁵ Such a norm is therefore justified. And this means we cannot avoid our problem by the strategy of ruling out testimonial evidence altogether.

Field suggests a more subtle strategy. Intuitively, if Matty had the courage of her convictions, and if she were ideally competent, her confidence in the proposition that T is a theorem of S would not be affected by the skepticism of her colleagues. If her proof is sound, she would see this clearly, and if her proof is not sound, she would not have been tempted by it in the first place. Field suggests, then, that we should ignore “computational limitations” in interpreting the empirical indefeasibility requirement.³⁶ However, a thinker with no computational limitations might nevertheless lack relevant concepts or lack some formal knowledge required in order to generate a given proof.³⁷ We need, therefore, to say something like this. Empirical considerations do not count as *empirical evidence* against a proposition if they would not undermine the credibility of the proposition to an “ideal thinker” – a thinker with no psychological

35 In some cases, we have reason to believe we are in a better epistemic position than others because we know that we are in an unusually good position.

36 Field 2000, 118.

37 Kitcher discusses similar issues (2000, 67–68).

weaknesses, with no computational limitations, and with a full conceptual repertoire.³⁸

Intuitively, after all, if her proof is sound, Matty can defeat the warrant-undermining effect of the skepticism of her colleagues on the a priori ground that she had to begin with, namely on the ground of the proof. Intuitively, the skepticism of her colleagues is perhaps reason for her to reconsider her proof, but it is not empirical evidence against her theorem. Intuitively, then, considerations that could themselves be defeated on a strongly a priori basis do not count as empirical evidence against a proposition.

5. ETHICAL NATURALISM AND THE STRONGLY A PRIORI

With all of this being clarified, I am finally in a position to explain why I think that moral naturalism is best understood in terms of the strongly a priori. To understand this, we need have before us a naturalistic picture of morality and of moral epistemology. I will attempt to sketch such a picture.

The underlying idea is that moral truths reflect empirical facts about human nature, the needs of societies, and the like. These facts vary from possible world to possible world, and so the moral truths also might vary from possible world to possible world. This is why we need to have experience of our world in order to have moral knowledge. Through our experiences early in life we come to have a particular moral perspective. We come to have moral concepts, both 'thick' and 'thin,' and, beyond that, we come to have a substantive normative 'theory' as to which things are right and wrong, good and bad, virtuous and vicious. Assuming that this perspective or theory is correct or approximately correct, we can come to be reliable in 'detecting' moral facts. We can also come to have (synthetic) moral beliefs that are warranted without the input of any empirical evidence beyond the experience that led to our initial acceptance of our basic moral perspective.

Consider cases of two kinds. First, it might be reasonable for us to believe that friendship is good, for example, even without empirical evidence that it is good. We might have acquired the concept of goodness while our parents were vainly encouraging us to form friendships, with the result that we came to think of friendship as good. In such a case, even

38 We might need to add that the beliefs of an ideal thinker are logically consistent and closed under entailment.

if we have not experienced any friendships, we might be reasonable to believe friendship is good. In this way, it might be 'default reasonable' for those in our culture to believe that friendship is good. Second, there can be cases in which we come to have moral beliefs after reflection, where our beliefs are warranted only in light of such reflection. Perhaps, for example, it is only after careful thought that we come to accept that there is no morally relevant difference between doing harm and allowing harm. In this case too, our belief might qualify as reasonable even though it is not based on empirical evidence beyond the experience that led to our initial acceptance of our basic moral perspective. In both kinds of case, the proposition we believe is such that it is reasonably believed by us, given our moral culture, without empirical evidence beyond the experience that led to our initial acceptance of our moral perspective. Such propositions are therefore weakly a priori even though it is perhaps only someone who shares our perspective who can reasonably believe these propositions without empirical evidence. The naturalist has no reason to deny the possibility of weakly a priori moral propositions of these kinds, for, of course, she would view our perspective as empirical in the sense at least that it could be undermined by empirical evidence.

The discussion already suggests that ethical naturalists can agree with nonnaturalists that we can have noninferential moral knowledge. That is, we can have knowledge of synthetic moral truths that is not based in any overt or conscious inferential reasoning. The dispute between naturalists and nonnaturalists therefore is not over the truth of *intuitionism*, at least not if we understand intuitionism to be the doctrine that we can have noninferential moral knowledge or warranted belief.³⁹ For example, a naturalist can agree that we might have a noninferential warranted belief that friendship is good. In some cases, our warrant might qualify as noninferential even if it depends on reflection, if, as Robert Audi has argued, the relevant kind of reflection is best seen as involving a response to a set of considerations rather than an inference from premises.⁴⁰ For example, our belief that there is no morally relevant difference between doing harm and allowing harm might qualify as noninferential in the relevant sense.

What has emerged, then, is that to distinguish ethical naturalism from nonnaturalism in the way I am proposing, on the basis of a distinction between the empirical and the a priori, we need the strong reading of the

39 This characterization of intuitionism is from Audi 1998, 19. Audi agrees that naturalism is compatible with the existence of intuitive moral judgment (25–28).

40 Audi 1998, 19–20 and 22–23.

a priori. We need to construe the moral naturalist as holding that moral properties are natural* in the sense that no synthetic propositions about their instantiation are strongly a priori. That is, no such proposition is such that both it can be reasonably believed without empirical evidence and it admits no empirical evidence against it.

6. WHY NATURALISM?

The upshot, on my account, is that the naturalist is committed to denying that there are strongly a priori synthetic moral truths. Now, then, we face the question, Why naturalism? Here I will discuss only one argument, an argument from epistemic defeaters. The argument is not decisive, but I think it captures an underlying motivation for naturalism, and it addresses the central issue.

We can narrow down the issue. The most plausible candidates for synthetic strong a priori status are moral generalities, such as the proposition that slavery is unjust, which I assume to be synthetic, or the Kantian thesis that we ought to treat humanity as an end in itself. For example, it would not be plausible to think it is strongly a priori that the slavery practiced in the 1850s in the United States was unjust unless it is plausible to think it strongly a priori that slavery as such is unjust. Otherwise, evidence about the circumstances in the 1850s could in principle undermine the reasonableness of our belief about the injustice of slavery in the United States. And I think the status of other specific moral judgments is relevantly similar. Hence, I think, the nub of the dispute between naturalists and nonnaturalists concerns the status of synthetic moral generalities, such as the proposition that slavery is unjust or the Kantian thesis I mentioned.

This brings us to the argument from empirical defeaters, which has three premises. First, no strongly a priori proposition admits of empirical evidence against it; any putative evidence against such a proposition would fail to undermine its credibility to an ideal thinker. Second, any synthetic moral generality *M* is such that there are possible experiences that, if they were actual, would at least *prima facie* constitute empirical evidence against *M*. And third, the undermining effect of such experiences on the credibility of *M* for a thinker need not be due to psychological weaknesses or computational limitations or to the thinker's lacking a full conceptual repertoire. It need not be due to the thinker's being less than ideal. It follows, the naturalist claims, that there are no strongly a priori synthetic moral propositions. And this is the central thesis of ethical naturalism as I understand it.

I have already explained the first premise in the argument and I have explained why I want to understand the naturalist as denying that there are any strongly a priori synthetic moral propositions. I turn now to the second premise, the claim that any synthetic moral proposition is such that there are possible experiences that, if they were actual, would at least *prima facie* constitute empirical evidence against it. The main argument for the premise is an argument from disagreement.

We should note in passing that particularists seem to be committed to the premise. Particularists hold that any general claim about the extension of a moral property can in principle be undermined by further experience.⁴¹ An example is the proposition that lying is morally wrong, which I assume to be synthetic. Particularists would argue that no such proposition is true, unless it is multiply qualified. For example, a particularist presumably would argue that there can be circumstances in which lying is permissible and that there can even be circumstances in which the fact that an action would be a lie might count in its favor. It appears, moreover, that particularists should agree that synthetic moral generalities of this kind are empirically defeasible, since experience can provide evidence that they are not true. And this means that they should agree that such propositions are not strongly a priori unless they would argue that the considerations that undermine them do not strictly speaking qualify as *empirical evidence* against them.⁴² But then, to avoid naturalism, particularists must rest their case on subtleties about the notion of empirical evidence. And they would need to argue either that certain synthetic moral generalities – propositions that they view as false – are nevertheless strongly a priori or that particular moral judgments can be strongly a priori. I have already suggested that this latter claim is implausible. Of course, the idea that particularism would push us in the direction of naturalism goes against conventional wisdom, and particularism is controversial. Let me therefore turn to the argument from disagreement.

The key idea is that our warrant for believing a proposition can be undermined or weakened by the disagreement of others in cases in which we have no independent reason to think we are in a better epistemic

41 For a discussion of particularism, see Dancy 2001.

42 A particularist might argue that the mere fact that evidence *could* crop up is sufficient to show it is not the case that lying *as such* is wrong; it is not necessary that the evidence *actually* crop up. The particularist might then claim that considerations that would count as *empirical* counter-evidence must actually obtain in order to do their work. This objection raises issues about a variety of matters that go well beyond the scope of this chapter, including issues about thought-experiments in science.

position than they are. I argued earlier in support of this thesis. If it is correct, then moral disagreement can weaken our warrant for our moral beliefs, and since disagreement is an empirical phenomenon, this supports the naturalist's thesis that synthetic moral generalities are not strongly *a priori*. Moral disagreement qualifies as empirical counter-evidence against our moral beliefs – unless such disagreement would not undermine the credibility of the beliefs to an ideal thinker.

The thesis that moral disagreement qualifies as empirical counter-evidence will seem plausible to the naturalist, of course. For she will say that since human nature and the needs of societies are, by and large, constant in our world, the moral beliefs of other people, when we cannot identify an error in their thinking and when we have no independent reason to think we are in a better epistemic position than they are, have a bearing on the reasonableness of our own moral beliefs. The naturalist will view people as detectors of moral facts, and she will view the disagreement of people who seem to be just as well placed as we are to detect the moral facts as evidence against our own beliefs.

Suppose, for example, that I witness a bullfight and observe that many thousands of people who seem to be good-hearted and fair-minded see nothing wrong in the treatment of the bull that takes place. As a result, I might begin to doubt that bullfighting is wrong, despite the 'harsh treatment' of bulls that is involved in bullfighting. But whether or not I begin to have doubts, if I cannot justify on independent grounds the claim that I am better placed epistemically to judge bullfighting than the people who attend bullfights, then the fact that so many people disagree with me about the wrongness of bullfighting would appear to constitute evidence against my belief that bullfighting is wrong, undermining or weakening my warrant for the belief. If this is correct, then the proposition that bullfighting is wrong is not strongly *a priori* – unless the undermining effect of the disagreement on the credibility of the proposition is due to psychological weaknesses or computational limitations or to the lack of a full conceptual repertoire such that the disagreement would not undermine the credibility of the proposition to an ideal thinker.

One might object that, even though I am aware that many decent people disagree with me, I would not be guilty of any kind of epistemic fault if I were to continue to believe that bullfighting is wrong. Disagreement is a two-way street. If I would be epistemically at fault to continue to believe that bullfighting is wrong, given the disagreement of the fans of bullfighting, then, by parity of reasoning, they would be epistemically at fault to continue to believe that bullfighting is permissible, given the

disagreement of the opponents of bullfighting. But if they are epistemically at fault to believe this, then how could the fact that they believe it mean that I am epistemically at fault to believe what I believe? It is important, however, to distinguish between the issue whether a belief is warranted and the issue whether a person is epistemically at fault to have it. The claim I am defending is that my warrant for believing that bullfighting is wrong is weakened by the disagreement. It is a separate issue whether I am epistemically at fault – perhaps, given my background, I am psychologically unable to take seriously the idea that bullfighting is permissible.

It is important to understand that it is no part of my argument that the possibility of well-meaning disagreement shows all synthetic moral generalities to be *false*. Moreover, I am not saying that the mere *possibility* of such disagreement is sufficient to show that I am not *actually* warranted to believe any moral generalities. The claim is rather that if there *were* to be disagreement of the kind I am imagining, and if I were aware of it, my epistemic warrant *would* be undermined.

We can provide examples involving disagreement about more basic moral principles than the thesis that bullfighting is wrong. Consider the Kantian thesis that we ought to treat humanity as an end in itself. Suppose one were to meet a group of well-meaning people, the Insiders, whose culture is such that although they know that those from outside their society are human, they view themselves as owing nothing to outsiders. The fact that the Insiders have this view would, I submit, undermine or weaken the reasonableness of believing that every human must be treated as an end in himself. It would be *prima facie* evidence against it, and so it would show that the proposition is not strongly *a priori* – unless, again, the disagreement would not undermine the credibility of the proposition to an ideal thinker.

To this point, the argument suggests that the reasonableness of believing a synthetic moral generality would be undermined by disagreement on the part of people who there is no independent reason to believe to be in a less favorable epistemic position than we are in. Of course, in any actual case of disagreement regarding a moral generality, it might be possible to show that we are actually in the better epistemic position for judging the truth than those who disagree with us. However, the argument is addressed to cases in which we have no independent reason to think that we are in the better epistemic position. The fact that some people disagree with us obviously is not an *independent* reason to think they are in a worse

epistemic position. It might be a reason to think they are morally worse than we are, but that is not the issue.

Naturalism does not yet follow from the argument. For it remains possible that there are certain moral generalities such that the fact that people disagreed with us regarding those generalities would not qualify as *empirical evidence* against them. Recall the idea we canvassed before, following a suggestion by Field, that putative evidence against a proposition is not genuinely *empirical* evidence if it would not affect the credibility of the proposition to an ideal thinker – a thinker who had no psychological weaknesses or computational limitations and who had a full conceptual repertoire. This brings us to the final premise in the argument from epistemic defeaters, the claim that in no relevant case of disagreement regarding a moral proposition M must the undermining effect of the disagreement on the credibility of M be due to psychological weaknesses or computational limitations or to the lack of a full conceptual repertoire, such that the disagreement would not undermine the credibility of M to an ideal thinker. If this premise is correct, then moral disagreement counts as empirical counter-evidence.

The premise is plausible on its face. For we are assuming that the moral generalities at issue are *synthetic*. It would not help nonnaturalism if we were to find an argument for a moral proposition M according to which any denial of M on the part of a thinker with an adequate grasp of all relevant moral concepts would have to issue from conceptual confusion. This would show M to be a conceptual truth, and there is no issue whether conceptual truths are a priori. The denial of a moral proposition that is a conceptual truth would not undermine its credibility to an ideal thinker, but this is not to the point. To be sure, there may be certain synthetic moral propositions that no *morally virtuous* person would give up, not even in the face of widespread well-meaning disagreement. There may be propositions that only a bad person or a morally confused person would deny. Indeed it seems to me that this is true.⁴³ Disagreement about such propositions would not undermine their credibility to a *morally* ideal thinker. But this also is not to the point. The ideal thinker of my argument is not guaranteed to be *morally* ideal.

I can see two ways in which one might attempt to block the argument. The first is to attempt to show, for certain synthetic moral propositions,

43 A morally good person would not deny that it is wrong to torture babies for fun. Tyler Burge urged me to consider such examples (in discussion).

that, despite the fact they are synthetic, any disagreement whether one of them is true must be due to the effect on one of the parties either of psychological weaknesses or computational limitations or an incompleteness in her conceptual repertoire. The second is to provide a strongly a priori argument for a synthetic moral proposition, an argument the premises of which can reasonably be accepted without empirical evidence and cannot be undermined by empirical evidence.

In pursuing the first strategy, one might invoke coherentism about epistemic warrant, for if the coherence or lack of coherence of a body of belief is an a priori matter, it might seem that it would be an a priori matter whether a person who disagrees with me about M is warranted. But what we need is a reason to think that any system of belief that includes the denial of M must be less coherent than it would be if it were changed minimally to include the acceptance of M. And this reason must be one that any ideal thinker could discern so that disagreement about M would not affect its credibility to an ideal thinker. But this means, in effect, that we need a reason to think there are certain *synthetic* moral generalities such that their denial is *incoherent*, even though it is compatible with the absence of any linguistic or conceptual confusion. I see no reason to think there are any such moral propositions.

The second way in which one might proceed would be to attempt to provide a strongly a priori argument for some synthetic moral generality. There are of course arguments, including Kantian arguments, that purport to show that we can have synthetic a priori moral knowledge or warranted belief.⁴⁴ I do not believe that any such argument has been successful. Unfortunately, however, I do not know how to prove this, and discussion of the arguments is clearly beyond the scope of this chapter. Here we reach rock bottom as far as this chapter is concerned.

To summarize, I have proposed that we take naturalism to be the view that moral properties are natural in the sense that no synthetic proposition about the instantiation of a moral property is strongly a priori. I have not proposed a naturalistic theory.⁴⁵ Instead I have offered the argument from epistemic defeaters. According to this argument, the fact that there could be *prima facie* empirical evidence against any synthetic moral claim means that such claims are not strongly a priori – unless the putative evidence would not undermine the credibility of the proposition to an ideal thinker. Of course, the argument I gave is not conclusive, but we at least see where

44 Kant 1785, part III.

45 I have offered such a theory in Copp 1995.

the central issue is to be found. If we are to be nonnaturalists, it seems, we must follow something like a Kantian strategy. We need an a priori argument for a synthetic moral truth.

7. CONCLUSION

My goal has been to lay the groundwork for the development of ethical naturalism, first, by explicating its central thesis, second, by locating the pivotal issue between naturalists and nonnaturalists, and, third, by giving some reasons to take naturalism seriously. I proposed that naturalism is best understood as the view that the moral properties are natural in the sense that they are empirical. I then pursued certain issues in the understanding of the empirical. The crux of the matter is whether any synthetic proposition about the instantiation of a moral property is strongly a priori in that it does not admit of empirical evidence against it. I think that the argument from epistemic defeaters undermines the plausibility of a priorism in ethics and supports the plausibility of naturalism.

REFERENCES

- Armstrong, David M. 1989. *Universals: An Opinionated Introduction*. Boulder, Colorado: Westview Press.
- Audi, Robert. 1998. "Moderate Intuitionism and the Epistemology of Moral Judgment." *Ethical Theory and Moral Practice* 1: 15–44.
- Boghossian, Paul, and Christopher Peacocke. 2000a. "Introduction." In Boghossian and Peacocke 2000b: 1–10.
- , eds. 2000b. *New Essays on the A Priori*. Oxford: Clarendon Press.
- Bonjour, Laurence. 1998. In *Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, N.Y.: Cornell University Press.
- Brink, David O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- Dancy, Jonathan. 2001. "Moral Particularism." *Stanford Encyclopedia of Philosophy* (Summer 2001 edition), ed. Edward N. Zalta. <http://plato.stanford.edu/archives/fall1999/entries/moral-particularism>.
- Darwall, Stephen, Allan Gibbard, and Peter Railton. 1992. "Toward *Fin de siècle* Ethics." *Philosophical Review* 101: 115–189.
- Field, Hartry. 2000. "Apriority as an Evaluative Notion." In Boghossian and Peacocke 2000b: 117–149.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.

- Goldman, Alvin I. 1994. "Naturalistic Epistemology and Reliabilism." *Midwest Studies in Philosophy* 19: 301–320.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.
- Harman, Gilbert. 1977. *The Nature of Morality*. New York: Oxford University Press.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Kant, Immanuel. 1785. *Grounding of the Metaphysics of Morals*. Trans. James W. Ellington. Rpt., Indianapolis, Ind.: Hackett, 1981.
- King, Jeffrey C. 1994. "Can Propositions Be Naturalistically Acceptable?" *Midwest Studies in Philosophy* 19: 53–75.
- Kitcher, Philip. 1983. *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.
- . 1992. "The Naturalists Return." *Philosophical Review* 101: 53–114.
- . 2000. "A Priori Knowledge Revisited." In Boghossian and Peacocke 2000b: 65–91.
- Kornblith, Hilary. 1994. "Naturalism: Both Metaphysical and Epistemological." *Midwest Studies in Philosophy* 19: 39–52.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Miller, Richard W. 1985. "Ways of Moral Learning." *Philosophical Review* 94: 502–550.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt., Cambridge: Cambridge University Press, 1993.
- . 1968. "Reply to My Critics." In *The Philosophy of G. E. Moore*, 3rd edition, ed. P. A. Schilpp, 535–687n. La Salle, Ill.: Open Court.
- . 1993a. "The Conception of Intrinsic Value." In *Principia Ethica*, ed. Thomas Baldwin, 280–298. Cambridge: Cambridge University Press.
- . 1993b. "Preface to the Second Edition." In *Principia Ethica*, ed. Thomas Baldwin, 1–37. Cambridge: Cambridge University Press.
- Papineau, David. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Quine, W. V. O. 1961. "Two Dogmas of Empiricism." In *From a Logical Point of View*, 2nd edition, 20–46. Cambridge, Mass.: Harvard University Press.
- Railton, Peter. 1989. "Naturalism and Prescriptivity." *Social Philosophy and Policy* 7: 151–174.
- Sturgeon, Nicholas L. 1984. "Moral Explanations." In *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, ed. David Copp and David Zimmerman, 49–78. Totowa, N.J.: Rowman and Allanheld.

Four Epistemological Challenges to Ethical Naturalism: Naturalized Epistemology and the First-Person Perspective

Ethical naturalism is the doctrine that moral properties, such as moral goodness, justice, rightness, and wrongness, are among the ‘natural’ properties that things can have. It is the doctrine that moral properties are ‘natural’ and that morality is in this sense an aspect of ‘nature.’ Accordingly, it is a view about the semantics and metaphysics of moral discourse. For example, a utilitarian naturalist might propose that wrongness is the property an action could have of being such as to undermine overall happiness, where happiness is taken to be a psychological property. Unfortunately, it is unclear what the naturalist means by a ‘natural’ property. For my purposes in this chapter, I shall assume that natural properties are such that our knowledge of them is fundamentally empirical, grounded in observation.¹ More precisely, a property is ‘natural’ just in case any synthetic

I am grateful for helpful comments from Richmond Campbell, Janice Dowell, Bruce Hunter, Loren Lomasky, Marina Oshana, Dave Schmidtz, Walter Sinnott-Armstrong, David Sobel, and Sara Worley. Abridged versions of this chapter were presented to the departments of philosophy at the Ohio State University, the University of Miami, and York University as well as to the 2000 Congress of the Canadian Philosophical Association. I am grateful for the suggestions and comments I received on these occasions.

- 1 G. E. Moore suggested a similar account when he said that naturalistic ethics holds that “Ethics is an empirical or positive science: its conclusions could all be established by means of empirical observation and induction” (Moore 1903, sec. 25). Naturalism is often explained differently, however, in terms of the sciences, or in terms of the entities postulated in the sciences or in certain favored sciences, or in terms of the vocabulary of the sciences or of the favored sciences. See, for example, King 1994, 53–56. Moore writes, “By ‘nature,’ then, I do mean and have meant that which is the subject-matter of the natural sciences and also of psychology” (1903, sec. 26). Ethical naturalism might then be described as the view that moral properties (and relations) are among the entities postulated in (the favored) scientific theories. Alternatively, it might be described as holding that moral terms can be ‘analyzed’ using the vocabulary of the (favored) sciences. The proposal I make in the text might lead a

proposition about its instantiation can be known only a posteriori, or with the aid of experience.² Ethical naturalism is, in short, the doctrine that there are moral properties and that they are natural properties. It implies that moral knowledge is fundamentally empirical. It is committed to a broadly ‘empiricist’ moral epistemology.

This chapter springs from the fact that certain unsurprising common-sense first-personal observations about our moral thinking can appear to undermine ethical naturalism by undermining the plausibility of the idea that our moral knowledge is empirical. For instance, moral belief seems more often to be a result of thinking an issue through in a way that is sensitive to morally significant considerations than a result of straightforward empirical observation or theorizing. I shall discuss four challenges of this kind. These challenges might be thought to support a kind of nonnaturalistic intuitionism in moral epistemology, or perhaps to support a kind of apriorism. I shall argue that the

philosopher to embrace (what we could call) ‘scientific naturalism’ of one these kinds if he holds, as many do, that the scientific method is privileged as a way of acquiring knowledge a posteriori. I do not see why philosophers would give a privileged place to the sciences in explicating naturalism unless they thought that the scientific method was at least an especially reliable way of acquiring knowledge a posteriori. Notice that, on the proposal made in the text, a naturalistic theory need not be ‘reductive.’ Notice also that a theory that is putatively naturalistic can be unsuccessful in a variety of ways. It might fail to make good on the claim that moral propositions are knowable, or that they are knowable a posteriori; it might propose an implausible analysis of moral propositions, or an analysis such that it is implausible that moral propositions are knowable or knowable a posteriori. Consider, for example, a divine command moral theory combined with a ‘natural theology.’ This is a form of moral naturalism, but an unsuccessful one, or so I believe.

- 2 More would obviously need to be said in order to give an adequate account of the a posteriori. By a “synthetic proposition” I mean a proposition that is neither a logical truth nor a conceptual truth. The idea is that a naturalist is one who denies that there is synthetic a priori moral knowledge. If, for example, the concept of murder is the concept of a wrongful killing, a naturalist would not deny that we can know a priori that murder is wrong. But a naturalist would deny that we can know a priori that, say, killing the innocent is wrong. The proposition that killing the innocent is wrong is synthetic, and in the relevant sense, it is ‘about the instantiation’ of wrongness since it implies that wrongness is instantiated in any action that is a killing of an innocent. The notion of a proposition ‘about the instantiation’ of a property is vague. Clearly, if a proposition implies that a property is instantiated, it counts as being ‘about the instantiation’ of the property. A proposition is also in the relevant sense ‘about the instantiation’ of a property if it implies a proposition about the circumstances in which the property would be instantiated. Consider, for example, the proposition that friendship is good. It is ‘about the instantiation’ of goodness since it implies that if there is friendship, it is good. G. E. Moore therefore counts as a nonnaturalist. He holds that we can know a priori that friendship is good. See Moore 25, secs. 112–113. But he also holds that the proposition that friendship is good is synthetic (sec. 6).

observations that fuel the challenges are actually compatible with ethical naturalism. In the process I will be defending the idea that moral ‘intuition,’ or noninferential spontaneous moral belief, can qualify as knowledge.

Some naturalists might be prepared to adopt the quite different strategy of disregarding objections of the kinds I will discuss on the basis of the metaphysical attractiveness of ethical naturalism. But such a strategy is not compatible with a naturalized approach to epistemology. As we will see, naturalized epistemology is undergirded by a kind of ‘scientific prioritarianism.’ Because of this, it blocks dismissing the objections on metaphysical grounds alone. Scientific prioritarianism appears also to block other kinds of responses to the objections, such as postulating a special faculty by which we acquire moral knowledge, or inferring how the psychology must work on the basis of metaphysical arguments. As we will see, naturalized epistemology gives priority to *scientific* psychology rather than to commonsense psychological observations of the kinds that fuel the objections. Hence, it would be compatible with naturalized epistemology to reject the objections if the ‘observations’ that fuel them are empirically suspect. It might also be compatible with naturalized epistemology to reject them on grounds of theoretical simplicity and explanatory utility. The important point, however, is that naturalized epistemology restricts the strategies that can be used by ethical naturalists in responding to the epistemological objections. To be sure, as we will see, ethical naturalism is not *logically* committed to the doctrines of naturalized epistemology. I nevertheless find it difficult to see how a theoretical preference for ethical naturalism could be explained or justified in a way that would not equally well ground or justify a theoretical preference for at least the central doctrines of naturalized epistemology.

There are two projects for the chapter. The most important is to respond to the intuitive epistemological objections to ethical naturalism. The second is to explain the relation between ethical naturalism and naturalized epistemology. Naturalized epistemology puts an important constraint on metaethical theory, namely, that its semantics and metaphysics must be integrated with a psychologically plausible moral epistemology. The four objections to ethical naturalism that I will discuss are grounded in an application of this constraint. On a naturalized approach to epistemology, an ethical naturalist cannot deal adequately with the objections without developing a moral epistemology that is both naturalistic, in that it shows moral knowledge to be fundamentally empirical, and compatible

with a psychology of moral belief formation and moral reasoning that is plausible by the standards of psychology and the (other) sciences. I sketch such an epistemology in this chapter. I will begin by presenting the objections in detail.

1. FOUR EPISTEMOLOGICAL CHALLENGES TO ETHICAL NATURALISM

According to the first objection, we often seem to arrive at our moral views as a result of reflection, thought, or reasoning, rather than as a straightforward result of empirical observation or theorizing, as naturalism would seem to suggest. Observation gives us information that is morally relevant. But we can be morally perplexed, say, about euthanasia, even if we are clear that no further observation will help us to decide what to think. In such cases reflection is called for rather than empirical theorizing about the world. Naturalism owes us an account of the nature and epistemic status of the relevant kind of reasoning or reflection and of how it gives us access to the empirical truths that it identifies with moral truths. It needs to explain how moral reasoning of this kind can give rise to knowledge if, as naturalism maintains, the basic moral facts can only be known empirically or through observation.³

Second, in many cases where we draw moral conclusions as a result of conscious inferences from observations, it would on the face of it be misleading to view these as inductive inferences of any kind, such as inferences to the best explanation, although this is what might be suggested by naturalism. For instance, once we determine that an act is, say, a piece of deliberate cruelty, such as an instance of torturing just for fun, it would be appropriate for us to conclude straightaway that the act is wrong. Yet the inference to the wrongness of the action clearly is not an inference to the best explanation of the fact that the act is an instance of torturing just for fun. And it would be misleading to describe the fact that the act is an instance of torturing just for fun as *evidence* that the act is wrong. Its support for the wrongness of the act is rather stronger than and different from mere evidence of wrongness.

3 The objection is briefly sketched by Scanlon (1998, 1). See also Audi 1996, esp. 114–115. For brief discussion of a similar objection, see Railton 1986, 166–168.

Naturalism owes us an account of the inferences we make in such cases.⁴

Third, although in some cases we do arrive at a moral belief directly on the basis of observation, without conscious inference, it seems inaccurate to view us in these cases as observing, say, the wrongness of an action. Naturalism would seem to suggest that we do or can observe the wrongness of actions in such cases, just as we might observe the clumsiness of someone's action. To see the problem with this suggestion, consider a variation on a well-known example that was introduced by Gilbert Harman.⁵ Suppose a person comes round a corner where some children are lighting a cat on fire in plain view, but suppose she does not see that what they are doing is lighting a cat on fire. Her failure might be explained by a fault in her perceptual apparatus, or perhaps by a lack of knowledge of cats. Perhaps she does not recognize that the animal being lit on fire is a cat. But suppose that although she sees and understands that the children are lighting a cat on fire, she does not 'see' that what they are doing is wrong. This is no evidence at all of a fault in her perceptual faculties, nor is it good evidence that she is lacking some propositional knowledge that she need only acquire to see things rightly. It is much better evidence of a fault in her moral sensitivity. Naturalism owes us an explanation of such cases and of the nature and epistemic role of moral sensitivity.

Finally, a naturalistic theory will likely be embarrassed if it proposes informative naturalistic accounts of the moral properties. For it seems likely that there will be cases in which we take ourselves to have moral knowledge even though we have no knowledge of whether the theory's proposed naturalistic explanans of what we know obtains or not. We might have no inkling of what the proposed explanans is. And it seems likely that, in attempting to decide what to believe in a case where we are morally perplexed, we will not investigate whether the naturalistic explanans obtains, but will rather engage in a more standard kind of moral reflection. For instance, we might be morally perplexed about euthanasia. A proposed naturalistic account of the proposition that, say, euthanasia is wrong, would be a general proposition about euthanasia that we could state in purely naturalistic terms. It might be the proposition that euthanasia undermines the general happiness; or the proposition that

4 An argument that assumptions about moral facts are irrelevant to explaining any observations is found in Harman 1977, ch. 1.

5 Ibid., 4–8.

a social rule against euthanasia would best serve the needs of our society, such as its need for peaceful social interaction; or the proposition that a rule that permitted euthanasia would be rejected by people who aimed to find principles for the general regulation of behavior that others with the same goal would not reject.⁶ If we were undecided whether euthanasia is wrong, it is unlikely that we would attempt to decide what to believe about it by attempting to decide whether to believe the empirical proposition about euthanasia that a naturalistic theory would identify as stating the truth conditions of the proposition that euthanasia is wrong. We would instead engage in ordinary moral reasoning of a familiar kind. We might express these points by saying that there is an apparent independence of moral belief from belief in the truth conditions of moral propositions that would be proposed by a naturalistic theory. Naturalistic theories need to explain this.

The four challenges form a cluster. There are cases in which we reach moral views as a result of reasoning or reflection. Naturalism owes us an account of what is going on. There are cases in which we infer a moral conclusion from an observation, but the inference does not seem to be inductive in nature. Naturalism owes us an account of what is going on in these cases. There are cases in which we come to have moral views immediately as a result of observation, but in these cases it seems it would be misleading to describe us as observing, for example, the wrongness of someone's act. And, finally, it seems not to be the case that we base our moral beliefs on knowledge of the complex empirical facts that a reductive naturalism would cite as constituting the truth conditions of these beliefs. In short, naturalistic metaethics does not seem to cohere with a plausible moral epistemology, given commonsense observations about moral belief and moral reasoning.

There is no need to accept naturalized epistemology to appreciate the force of these four challenges. But if we accept a naturalized epistemology, we are committed to certain restrictions on acceptable responses. Most important, we must concede that the philosophical soundness of our response is hostage to its psychological plausibility. In the [next section](#) of the chapter, I address the basis of this idea in naturalized epistemology. This section is optional for readers who are primarily interested in my responses to the four challenges.

6 In this sentence I allude to three naturalistic proposals: a form of analytic consequentialism, the view I proposed in Copp 1995, and a close relative of the view one can see in T. M. Scanlon's 1998 book (4). Scanlon does not intend to propose a form of ethical naturalism.

2. WHAT IS NATURALIZED EPISTEMOLOGY?

Philosophers who have discussed something called “naturalized” epistemology have had in mind different doctrines about epistemology, and they have disagreed about the plausibility of these doctrines. It is to be expected, then, that my understanding of naturalized epistemology is different from that of many other philosophers. Fortunately, this does not matter for my limited purposes. For my purposes, moreover, a brief discussion of naturalized epistemology will suffice.

Quine’s central concern, in his classic paper, “Epistemology Naturalized,”⁷ was the failure of “traditional epistemology” to deal with skepticism about the external world. In face of this failure, Quine recommended that epistemology give up the “Cartesian quest for certainty” and instead see itself as a part of psychology, exploring empirically the relation between evidence and theory. He says, “epistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject.” It studies “[t]he relation between the meager [sensory] input and the torrential output,” which comes in the form of a description of the natural world.⁸

I think we should abandon Quine’s radical idea that epistemology is rightly seen as a part of psychology. There are normative issues in epistemology that this view cannot accommodate, and we do not need to embrace the “Cartesian quest for certainty” in order to address these issues.⁹ Our choice is not the stark one that Quine poses between

7 Quine 1969a.

8 Ibid., 74–76, 82–83.

9 Here I agree with Barry Stroud (1985) and Jaegwon Kim (1993). Perhaps it will be objected that normative epistemological issues, properly understood, *are* psychological. On this way of thinking, we would perhaps need to rethink the import of Quine’s view that “Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science.” I take Quine to be claiming that we should give up normative epistemology as it has been practiced and instead confine ourselves to the use of scientific methodology in exploring the relation between evidence and theory. But the view that normative epistemological issues are psychological could be understood to imply instead that we should expand our conception of psychology, natural science, and the scientific method so that traditional philosophical explorations of normative issues are counted as ‘scientific’ or ‘psychological.’ This does not appear to be a substantive suggestion. Note: The thesis that all normative properties are natural properties does not imply that normative epistemology is “a chapter of psychology”; what it implies is that any normative epistemological knowledge is fundamentally empirical. In more recent work, Quine appears to accept that epistemology is a normative discipline. See Quine 1990.

Cartesianism and psychologism. But even if we give up Quine's psychologism, there remains in Quine a less radical conception of naturalized epistemology as an antiskeptical, or at least nonskeptical, empirically informed, investigation of the grounds of knowledge. I shall take this revised conception of naturalized epistemology as my starting point.

Suppose that we reject skepticism about the external world. In so doing, we take it to be possible for ourselves, as we actually are constituted, to have knowledge about the world around us. This idea commits us to allowing that the psychological processes by which we come to have beliefs about the external world could underwrite knowledge. It commits us, that is, to placing certain psychological constraints on our philosophical account of what is necessary in order for a belief to count as knowledge. If we hold that it is necessary for us to stand in a certain relation to the world in order to have knowledge of it, or if we hold that it is necessary for us to go through a certain process of justification in order to have knowledge, then we are committed to the possibility of our actually going through this process or standing in this relation to the world, given how we actually are constituted. We are committed to constraining our philosophical epistemology by what is psychologically possible for beings like us. If we take ourselves *actually* to have knowledge of the world around us, then we are committed to a stronger thesis. For if we take ourselves to know that there are oak trees and stars, for example, we are committed to thinking that the actual psychological processes by which we come to believe such things are processes that yield knowledge, at least in some cases.

As I understand it for my purposes in this chapter, then, 'naturalized epistemology' is characterized by two central doctrines. First, we do have knowledge of the world around us. The actual psychological processes by which we come to have the relevant beliefs about the world around us are sources of knowledge.¹⁰ Second, in attempting to explain how these processes enable us to stand in epistemically relevant relations

10 This does not mean that an epistemological naturalist would ignore or reject out of hand the traditional skeptical challenge to our knowledge of the external world. Quine appears to think that science can give us a kind of response to skepticism, for, as he pointed out, science can at least hope to explain why it is that our experience leads us to have largely correct beliefs. He says, "There is some encouragement in Darwin. . . . Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind" (Quine 1969b, 126). But an epistemological naturalist need not think that this point provides an adequate philosophical response to skepticism. A variety of views about skepticism are compatible with naturalized epistemology.

with the objects of our beliefs, our philosophical epistemology must be constrained by a plausible and empirically informed psychology of these processes.¹¹

We know in general terms what these processes are. Perhaps most of us learn that there are oak trees and stars in the course of learning the language. We learn that the Big Dipper points to the North Star in early star gazing, at least if we live in the Northern Hemisphere. In these examples we learn from experience. Quine said, “The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world.”¹² This dictum might need to be qualified, for we do not want to be too quick to rule out the possibility of a priori knowledge. But it is plausible that, “ultimately,” even our a priori knowledge, if any, is dependent on the stimulation of our sensory receptors.¹³

The picture I have painted so far does not tell us what would be involved in naturalizing *moral* epistemology, for it does not tell us what would be involved in naturalizing the epistemology of a specific subject matter. Let me turn again to Quine. He says that in investigating the epistemology of science, “we are well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand.”¹⁴ This reveals that Quine is assuming that the theories of the science we are seeking to understand can be taken as a given. For he would not have said this if he had had in mind the epistemology of a theory that we take to be false, such as astrology. We could ‘naturalize’ the epistemology of astrology, or of another false theory or discredited science, in the sense that we could study empirically how

11 I am grateful to Bruce Hunter for help with this paragraph. Hilary Kornblith distinguishes two questions about our beliefs. The first is, “How ought we to arrive at our beliefs?” The second is, “How do we arrive at our beliefs?” He says that “the naturalistic approach to epistemology consists in [the thesis that] question 1 cannot be answered independently of question 2. Questions about how we actually arrive at our beliefs are thus relevant to questions about how we ought to arrive at our beliefs.” On my understanding, epistemological naturalism accepts the thesis of Kornblith’s “naturalistic approach” but adds two additional doctrines. First is an explanation of *why* epistemological naturalism accepts this thesis. It does so because it is antiskeptical or at least nonskeptical. Second, naturalism does not merely hold that the psychology of belief formation is *relevant* to normative issues in epistemology. It holds that our normative epistemology is to be *constrained* by the psychology of belief formation. See Kornblith 1985a, 1, 3.

12 Quine 1969a, 75.

13 I shall leave open the question whether there is a priori knowledge, and if so, how it should be understood. Nothing in this chapter turns on our having an answer to the question.

14 Quine 1969a, 76.

people came to accept it. But in so doing we would not use “information” provided by the science or theory in question.

These remarks point the way, I think, to a proper understanding of what would be involved in ‘naturalizing’ the epistemology of a specific subject matter. Naturalized epistemology is characterized by a nonskeptical doctrine combined with a methodological doctrine to the effect that our philosophical epistemology must be constrained by a plausible psychology of the processes whereby we acquire knowledge. In studying the epistemology of a specific theory or body of beliefs, T, an epistemological naturalist would aim to arrive at corresponding specific doctrines, first, about the cognitive status of T, and second, about the cognitive status of the psychological processes that have led people to accept T. First, to simplify somewhat, the naturalist would have to decide whether or not to take T to be true. More realistically, her decision might be more nuanced. She might decide, for example, that some of the beliefs or some of the propositions in T are true and some are not. In many cases, the issues here will be philosophically subtle and controversial. Second, depending on what she decides about the cognitive status of T, she will reach corresponding or at least compatible conclusions about the cognitive status of the processes that have led people to accept T. If she takes T to be true or parts of T to be true, then, depending on her normative epistemology, she presumably would take at least some of the psychological processes by which we come to accept T as sources of knowledge. And if so, then as Quine suggests, she can use or assume T or the known parts of T in studying these processes. If she takes T not to be true, then she must allow that these processes are potential sources of error, and she cannot use T or assume T in studying the processes. In more nuanced cases, her views about the cognitive status of the mechanisms that have led to acceptance of T will need to be more subtle and nuanced. A plausible naturalized epistemology of astrology would be built on the premise that the fundamental doctrines of astrology are false and not known. But a plausible naturalized epistemology of mathematics would instead be built on the premise that the theorems of mathematics are known. A naturalist would presumably take a more nuanced view of theoretical physics, given that even the best theories in physics are sensibly taken to be open to revision.

In order to get the project of naturalizing *moral* epistemology off the ground, the naturalist must decide whether we have moral knowledge, and, if she decides that we do, she must decide what it is that we know, at least within broad limits. She has to decide the truth conditions of our moral beliefs. The problems here are the familiar problems of moral

philosophy. The main contribution made by the epistemological naturalist would be the idea that our theories of what it is that we know, in having moral knowledge, and our theories of how it is that we know these things, must mesh with a plausible psychology of moral belief. An acceptable moral semantics and metaphysics must fit with an acceptable moral epistemology and an acceptable moral epistemology must fit with an acceptable empirical psychology of moral belief.

This is a familiar subtext in recent moral theory. According to nonnaturalism, moral properties are not natural properties, so our knowledge of their instantiation is not empirical. It is widely agreed that nonnaturalism owes us an explanation of how we can come to know that such properties are exemplified. How can we come to know that torture is wrong, for instance, unless wrongness is a property we can be acquainted with or otherwise related to in the natural world? How can we come to be in epistemically significant relations to nonnatural properties given that, ultimately, all our knowledge is grounded in observation of the natural world? The nonnaturalist might posit a special faculty by which we can detect the wrongness of torture.¹⁵ But according to epistemological naturalism, such a view is tenable only if there are independent psychological grounds for supposing the existence of such a faculty – or no independent psychological grounds for supposing its nonexistence. According to the epistemological naturalist, our psychological holdings cannot properly be amended to serve the needs of our philosophical theories. Rather, our epistemology is properly constrained by the holdings of empirical psychology.

Of course, at any given time, the ‘holdings’ of psychology are open to revision. Science is fallible, and there is never going to be a ‘finished’ psychology. There are controversies within psychology. We might view some of the tenets of a certain psychological theory, as we find it at a given time, to be empirically or otherwise scientifically suspect. Because of this, the characteristic doctrine of naturalized epistemology needs to

15 The intuitionism proposed by Robert Audi does not postulate a special faculty (1996, 121, 124). J. L. Mackie challenged any theory that postulates the existence of moral properties to provide a plausible epistemology of those properties. If a theory postulates the existence of a property of wrongness, for example, Mackie challenges it to provide a plausible account of how it is that we ‘discern’ the wrongness of actions that we believe to be wrong, and of how it is that we ‘discern’ the link between the actions’ feature of wrongness and the natural features, such as deliberate cruelty, that we believe the wrongness of the actions to be ‘consequential’ to. He appears to think that no such account will be as comprehensible and as simple, and as plausible psychologically, as the idea that we are not perceiving a property of wrongness at all, but are rather simply responding subjectively and negatively to the natural features in question. See Mackie 1977, ch. 1, sec. 9.

be interpreted with care. It is too crude to hold that our epistemology is to be constrained by the ‘holdings’ of psychology. Instead, we should say this: If we take it that a given thesis is settled in current psychology and is not scientifically doubtful, then our epistemology must be compatible with that thesis. The doctrine is roughly that our epistemology must be constrained by what we take to be the settled results of empirical psychology regarding the formation of our beliefs.

This doctrine is not uncontroversial. It rules out amending the results of psychology in the interest of explaining the possibility of our having knowledge of a specific subject matter. Paul Benacerraf used an argument of this kind against Kurt Gödel’s mathematical intuitionism.¹⁶ The doctrine could also be used in arguments against analogous intuitionist or special faculty views in moral epistemology and in the epistemology of theological belief. But it is not clear what argument could be given in support of the doctrine. Suppose that certain philosophical theories in mathematical and moral epistemology and metaphysics are in tension with what the settled psychology of a given period tells us about mathematical and moral belief formation. The thesis that, in this case, it is the philosophical theories that must give way, not the psychology, is itself a philosophical thesis. It is presumably derived from a more general doctrine to the effect that philosophy must be constrained by the results of science.

16 Benacerraf argued that “the concept of mathematical truth . . . must fit into an over-all account of knowledge in a way that makes it intelligible how we have the mathematical knowledge that we have. An acceptable semantics for mathematics must fit an acceptable epistemology” (1973, 667). Benacerraf then went on to argue that the standard “platonistic” account of mathematics, according to which numbers are abstract objects, “makes it difficult to see how mathematical knowledge is possible,” given a familiar causal account of knowledge (673, 671–673). On any acceptable epistemology, he suggests, there must be some “link between our cognitive faculties and the objects known.” For “[w]e accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties.” To fill this gap, one might postulate the existence of a special cognitive faculty, which we could call “mathematical intuition.” Benacerraf notes that Kurt Gödel postulated the existence of just such a faculty to account for mathematical knowledge, and he comments that Gödel evidently disagrees with him about what psychological processes are available to an acceptable epistemology. Benacerraf concludes that “the absence of a coherent account of how our mathematical intuition is connected with the truth of mathematical propositions renders the over-all account unsatisfactory” (674–675). Benacerraf does not explicitly say that a satisfactory epistemology of mathematics would have to mesh with an empirically plausible psychology of mathematical belief. But the idea is implicit in his response to Gödel’s intuitionism. He appears to rule out the soundness of an argument from metaphysics and epistemology to the existence of a psychological faculty. He therefore appears to accept tenets that would be characteristic of a naturalized mathematical epistemology. For similar views, Benacerraf cites Steiner 1973.

We could call this doctrine “scientific prioritism,” since it gives methodological priority to science. Of course, at any given time, the ‘results’ of science are open to revision, for science is fallible, and it is never going to be ‘finished.’ With this understood, the doctrine of scientific prioritism should be understood to say, roughly, that our philosophical theorizing must be constrained by what we take to be the settled results of empirical science. It is not clear what arguments could be given in support of scientific prioritism, but it is nevertheless characteristic of naturalized epistemology as I understand it.¹⁷

I also believe that any ethical naturalist would be tempted by scientific prioritism, given what I take to be the underlying epistemological motivation of naturalism. But it is not the case that an ethical naturalist is logically committed to naturalized moral epistemology. Ethical naturalism is the view that moral properties are natural properties. The parallel view in epistemology is the view that normative epistemological properties are natural properties. An ethical naturalist is not even logically committed to this latter thesis, and, anyway, it is distinct from naturalized epistemology. Naturalized epistemology is a position about the methodology of epistemology rather than a view about the metaphysics of normative epistemological properties. Ethical naturalism and naturalized epistemology are therefore logically independent of one another.¹⁸ They do nevertheless appear to be intellectual cousins. So it is important to consider whether ethical naturalism can adequately respond to the four epistemological challenges I described earlier in the chapter without running afoul of scientific prioritism or any of the other tenets of naturalized epistemology.

Indeed, the four challenges can now be seen to be commonsense instances of a more general theoretical challenge to ethical naturalism from naturalized epistemology. The four challenges suggest that there is evidence in commonsense reflection and observation that the nature of moral reasoning, moral inference, and moral observation are not what

17 Richmond Campbell objected, in personal correspondence, that epistemology and psychology are interdependent, since psychological methodology reflects certain assumptions about epistemology, and since epistemological theory depends on certain assumptions about psychology. But I don’t see this interdependence as a decisive objection to scientific prioritism. It may still be true that epistemology must be constrained by what we take to be the settled results of empirical science.

18 Naturalized epistemology is not committed to ethical naturalism for it is not committed to the idea that moral knowledge is empirical. It is compatible with naturalized epistemology to hold that we do not have moral knowledge, or to hold that our moral knowledge is not empirical. Similarly, it is compatible with naturalized epistemology to deny that we have mathematical or theological knowledge, or to hold that such knowledge is not empirical.

we would expect if ethical naturalism were true. That is, the metaphysics and semantics of ethical naturalism seem not to cohere with common-sense observations about moral belief. Naturalized epistemology adds that although common sense is not decisive, empirical psychology (among other things) can be decisive in assessing the tenability of a moral epistemology. Any given naturalistic metaethical theory must show that the psychological mechanisms that *actually* account for our moral beliefs, in cases where the theory implies that the beliefs are true, and so might constitute knowledge, are such as to give us knowledge-enabling access to the facts that, according to the theory's semantics and metaphysics, make the beliefs in question true. More briefly, an acceptable moral epistemology must fit with the psychological facts about how we come to have moral beliefs.

3. SUMMARY OF A NATURALISTIC THEORY

Before we can attempt to deal with these epistemological challenges, we need to bring ethical naturalism into clearer focus. For, to deal with the challenges, we need to show that there is at least a defensible and coherent form of ethical naturalism, and an accompanying moral epistemology that meshes suitably with commonsense observations about moral reasoning, moral inference, and moral observation and with the psychology of moral belief formation. To show this, we need to specify a naturalistic theory. In this section of the chapter, I will sketch such a theory. Fortunately, it is not necessary that I try to establish its truth. It will be enough to formulate it and to urge that it is defensible. For obvious reasons, the theory I will present for consideration is a naturalistic theory that I have proposed and defended in detail elsewhere.¹⁹ I have called it “society-centered moral theory.”

To a first approximation, the central idea is that a basic moral *proposition* is true only if a corresponding moral *standard* or *norm* is relevantly justified or authoritative.²⁰ By a “standard,” I mean a content expressible by an

19 Copp 1995. For a brief introduction to the view, see Copp 1997. For a reply to some objections, see Copp 1998.

20 A basic moral proposition is such that, for some moral property M, it entails that something instantiates M. An example is the proposition that capital punishment is wrong. Among nonbasic moral propositions are propositions such as that nothing is morally wrong and that either abortion is wrong or $2 + 2 = 4$. In Copp 1995, I called basic moral propositions “paradigmatic.”

imperative. For example, it is wrong to torture just for fun just in case (roughly) a standard or rule prohibiting people from torturing just for fun is relevantly justified.²¹ A moral standard is relevantly justified just in case (roughly) its currency in the social code of the relevant society would best contribute to the society's ability to meet its needs – including its needs for physical continuity, internal harmony and cooperative interaction, and peaceful and cooperative relations with its neighbors. This semantics treats moral properties as relational. If torture is wrong, it is wrong in relation to a given society, a society in which the currency of a standard prohibiting torture would best contribute to the society's ability to meet its needs. The moral standards with currency in a society form the social moral code of the society; a social moral code is a system of moral standards or rules that has currency in a society. Not all possible codes of this kind are relevantly justified or authoritative, of course. Moral claims are true or false depending on the content of the relevantly justified and relevantly local moral code.

Society-centered theory raises a number of difficult questions, including questions about its semantics and its metaphysics of moral properties. There are perhaps two sets of issues that are especially pressing. First are issues raised by the theory's treatment of moral properties as relational to societies. What distinguishes societies from other kinds of collective entities? Which society is the relevant one for assessing the truth of a given moral claim? Which society is the one, the needs of which determine my duties, for instance? Second are issues raised by the idea that the truth value of a moral claim is determined by the nature of the moral code the currency of which would best enable a society to meet its needs. What in detail are the needs of societies? Is there in general a *single* code the currency of which would *best* enable a society to meet its needs? I have addressed many of these questions in previous writings and I have introduced clarifications, qualifications, and amendments to the basic idea of the theory in order to deal with them.²² This is not the place to go into detail, or to attempt to explain why I find the theory plausible. Some of the details are less important than others for my purposes. In this

21 In a fuller discussion of society-centered theory, I would qualify this claim. It is correct that a basic moral claim is true only if a corresponding standard is relevantly justified, but there are other conditions necessary for the truth of some moral claims. Some details are set out in Copp 1995, e.g., 24–26, 28–30.

22 Copp 1995, chs. 6–11; Copp 1997.

chapter, I merely want to let the theory serve as an example of ethical naturalism.

According to society-centered theory, moral properties relate actions, persons, traits of character, institutions, and the like to the requirements of the moral code that is relevantly justified in relation to a relevant society. For my purposes, the important point is that moral properties are natural properties according to the theory. I proposed before that a property is natural just in case our knowledge of it is fundamentally grounded in observation and inference from observation. More accurately, a property is 'natural' just in case any synthetic proposition about its instantiation can be known only a posteriori, or with the aid of experience.²³ So in order to show that moral properties are natural according to society-centered theory, I need to show that our knowledge of them is fundamentally empirical.

Consider the property of wrongness, for example. To show that wrongness is a natural property, according to the theory, I need to show that, according to the theory, any knowledge we have about its instantiation is fundamentally empirical. Ignoring certain qualifications that are irrelevant here, the theory implies that the property of wrongness – in relation to society S – is the property of being forbidden by the social moral code the currency of which in society S actually would best enable S to meet its needs. Call this the "S-ideal moral code." That is, roughly, in a context in which society S is the morally relevant one, the term "wrong" picks out the property that would also be picked out by the complex description, "the property of being forbidden by the S-ideal moral code." Call this property "society-centered wrongness," or "SC-wrongness." I take it to be obvious that we could not know anything substantive about *this* property, such as whether or not some action has the property, except with the aid of experience. Except with the aid of experience, for example, we could not know that capital punishment has the property of being SC-wrong. We surely could not know a priori that capital punishment has the property, for some society S, of being forbidden by the moral code the currency of which in S actually would best enable S to meet its needs.

According to society-centered theory, every moral property is identical to some relation between things that have the property and the

23 The issue is what human beings could know, given their nature. Perhaps a god could know a priori things that humans could know only through experience. If some humans can know certain things a priori, then these things are knowable a priori regardless of whether some other humans would need to rely on experience in order to know them.

requirements of the S-ideal moral code. I have used the term “SC-wrongness” to pick out the property that the theory implies to be identical to wrongness. We can understand the terms “SC-virtue,” “SC-justice,” “SC-rightness,” and so on, in corresponding ways. On this account, for instance, the SC-virtues – in relation to society S – are, roughly, the states of character that all adults would be enjoined to exhibit by the S-ideal moral code. Since every moral property is like wrongness in the relevant way, every moral property is natural.

Given what I have said so far about society-centered theory, the theory might seem to imply that, in order to have any moral knowledge, we would have to be virtuoso sociologists. For example, to know that capital punishment has the property of being SC-wrong in relation to society S, it will seem that we would have to know a great deal about S and its circumstances so that we could know what a moral code would have to prohibit in order to best serve as the social moral code in S. So understood, society-centered theory will seem to be open to the epistemological challenges that we looked at before, as I will now explain. This should be no surprise. Naturalism attracts the challenges because it claims that moral properties are natural, and society-centered theory, as a kind of naturalism, makes precisely this claim. With society-centered theory on the table, then, let us go through the challenges one by one and ask ourselves how a defender of the theory might respond, compatibly with the constraints of naturalized epistemology.

4. MORAL BELIEF AND NATURALISTIC TRUTH CONDITIONS

I begin with the fourth challenge. According to this challenge, if a naturalistic theory claims to provide informative reductive accounts of the truth conditions of moral propositions, there are the following two problems. First, there will be cases in which we take ourselves to have knowledge that p, for some moral proposition p held by the theory to have truth conditions q, even though we do not believe that q, and even though we have no idea that q might express the truth conditions of p. Second, even in cases in which we arrive at a moral belief p as a result of deliberation, where p is held by the theory to have truth conditions q, it typically is not the case that we base our belief that p on knowledge that q, or even on a belief or evidence that q.²⁴

24 Here I take a ‘fine-grained’ view about the individuation of propositions. On a different view, I would need to reformulate what I say.

According to society-centered theory, for example, capital punishment is wrong just in case it is SC-wrong. And this is true just in case it has the property, roughly, of being forbidden by the S-ideal moral code. The latter property is quite a complex empirical property. It appears that we might believe that capital punishment is wrong without believing that capital punishment has this complex property. And it appears that people typically do not base their beliefs about the wrongness of capital punishment on knowledge of or evidence of such a complex fact about capital punishment as whether it is SC-wrong. To be sure, people sometimes argue that capital punishment is wrong on the ground that it does not deter murder, and this claim is at least *relevant* to the issue of whether the currency of a prohibition of capital punishment in the social moral code would affect the ability of a society to meet its need for internal harmony. But the connection to issues about the ability of society to meet its needs is not often drawn. And many people who are morally opposed to capital punishment at bottom simply view it as abhorrent for the state to take someone's life when that person is in custody and therefore poses very little threat to anyone. Similarly, many people who take capital punishment not to be wrong have at bottom a rather visceral belief that a person who is guilty of murder deserves a similar fate. There appears, then, to be a lack of connection between the grounds of moral belief and the truth conditions of moral propositions that are proposed by society-centered theory.

This is no objection in cases in which it is plausible that a person's moral beliefs do not qualify as knowledge. Suppose for example that someone believes capital punishment is wrong on superstitious grounds. Suppose the ouija board told him that capital punishment is wrong. In this case, we would not be tempted to think that the person knows capital punishment to be wrong, and so it is no objection to point out that his belief is not grounded in any belief that, by the lights of society-centered theory, would support the truth of the proposition that capital punishment is wrong.

The sharp end of the objection turns on the claim that our moral beliefs can *qualify as knowledge* in many cases where we do not believe, or have evidence or reason to believe, that the proposed society-centered truth conditions of our beliefs obtain. This seems to imply that the truth conditions proposed by society-centered theory are not correct. How could it be that moral propositions have the truth conditions that society-centered theory claims them to have if we can know a moral proposition to be true without having any evidence that, or believing that, or even

having any inkling whether the theory's proposed truth conditions for the proposition obtain?²⁵

A defender of the society-centered view might reply that a moral belief does *not* qualify as knowledge *unless* it is grounded in evidence that the truth conditions of the proposition obtain according to the account given in the theory. This would mean that moral knowledge requires detailed sociological evidence regarding the content of the S-ideal moral code. The theory certainly suggests that we *can* acquire moral knowledge in this way, at least in principle. I myself have presented armchair sociological arguments in an effort to ground some specific moral judgments directly in society-centered theory.²⁶ But we rarely deliberate about whether or not the proposed society-centered truth conditions of moral propositions obtain, and it appears that we rarely have any inkling of whether the proposed truth conditions obtain. I think that we have moral knowledge in many more than these rare cases.²⁷ How could this be, if society-centered theory is correct?

This is a serious challenge, and I think that a corresponding challenge could be developed for any naturalist theory that proposed substantive truth conditions for moral propositions. I now want to argue that the challenge can be met. To do so, I need to show, or at least to make plausible, that moral belief is relevantly analogous to beliefs in other areas where we observe similar phenomena.

Consider our 'economic beliefs,' which have complex truth conditions analogous to the truth conditions attributed to moral beliefs by society-centered theory. Ordinary people would have difficulty explaining the truth conditions of their economic beliefs in any substantive way. Consider the proposition that I have a U.S. one dollar bill in my hand. On reflection, I hope it will be obvious that the truth conditions of this proposition are enormously complex. In order for this piece of paper to be a one dollar bill, there must be a complex disposition among the relevant officials and among the bulk of the population of the United States to exchange the piece of paper for goods or services priced at one dollar. But what must obtain in order for the given population to be the population of the United States? What must obtain in order for goods or services to

25 I take it that this objection is an epistemic variation on the famous "open question argument." See Moore 1903, sec. 13.

26 See the discussion of moral issues, including abortion and cruelty to animals, in Copp 1995, 201–209, 213–216.

27 See *ibid.*, 237–240. See also Copp 1996.

be ‘priced’ at ‘one dollar’? What must obtain in order for a person to be one of the ‘relevant officials’? Plainly, we can know immediately, on inspection, that the thing in my hand is a U.S. one dollar bill, and we can know this without having more than a vague idea of the truth conditions of the proposition that it is a one dollar bill. Indeed, one dollar bills could not fulfill their function in the economy if this were not the case.²⁸ Our beliefs about the presence of dollar bills in our immediate visual field are reasonably reliable even though, remarkably, the property of being a one dollar bill is a highly complex theoretical property in a theory that few of us know.

Given the example of the one dollar bill, it seems clear that we can know something even if there is a proposition that expresses its truth conditions and we do not know that it is true.²⁹ The example did not involve deliberation, however, and one might think that this is significant. But one can know something as a result of deliberation while having no inkling about its truth conditions. If you see that I have both a one dollar bill and a five dollar bill in my hand, you can perform a simple mathematical calculation and conclude that I have at least six dollars without knowing the truth conditions of what you know as they are laid out in economics. This is no objection to the economic theory of the truth conditions of propositions about money. Perhaps, then, it is no objection to a metaethical theory that we can have moral knowledge without having any idea of the truth conditions of our moral beliefs as they are laid out by the theory.

The familiar ‘reliabilist’ strategy for responding to skepticism about the external world is basically to argue that we can have knowledge about the physical objects around us in circumstances in which we are detecting the objects by means of a reliable belief-generating mechanism. We can have such knowledge even if we cannot justify our beliefs about the objects by inferring them from data given in our sensory experience.³⁰ This strategy implies that knowledge is possible in cases in which we do not

28 This was pointed out to me by Elijah Millgram. If one dollar bills were counterfeited very commonly, then matters would be different. In that case, the one dollar bill would not serve very well its intended function in the economy.

29 To give a second example, it is clear that we can know various propositions about water without knowing, believing, or having any idea of the chemical truth conditions of the propositions. And the grounds of our ‘water beliefs’ need not include any grounds to believe corresponding propositions about H_2O that constitute the truth conditions of our water beliefs. But see above, note 24.

30 One of the first to propose and defend ‘reliabilism’ was Alvin Goldman (1979).

have knowledge of relevant truth conditions. But it does not follow that knowledge is possible without knowledge of truth conditions in a case in which the relevant belief is *not* produced by a reliable belief-generating mechanism.

Reliabilism therefore suggests a problem for our response to the objection about knowledge of truth conditions. We might concede that it is possible to know a proposition to be true without having any inkling of its truth conditions in cases in which the relevant belief is produced by a reliable belief-generating mechanism. We are reasonably reliable, in at least a wide range of circumstances, in forming true beliefs about dollar bills. But, one might say, given the moral disagreement we see in the world, it is not plausible that there is a similarly reliable psychological process for generating moral beliefs. If this is correct, then to defend society-centered theory, either we must argue that there *is* a reliable mechanism for generating moral beliefs, or we must after all deny that moral knowledge is possible without evidence as to the obtaining of the proposed society-centered truth conditions.

The best response to this objection, I think, is to argue that, at least in certain contexts, some people can be quite reliable in arriving spontaneously at true moral beliefs, and, moreover, their reliability in these contexts can be explained in a way that is compatible with naturalism. This is what I shall now argue.

5. MORAL SENSITIVITY

In Harman's example, we imagine ourselves coming around a corner and seeing some children lighting a cat on fire right in front of us.³¹ If we recognized what was going on, we naturally would know right away that what the children were doing was wrong. In contexts of this kind, we arrive immediately at a moral belief, and our belief might qualify as knowledge – assuming we are correct in our perception of what is going on. We might have no more than a vague and unhelpful idea of the complex truth conditions that would be assigned to the belief by society-centered theory. The example of the dollar bill suggests that this is no objection to the theory. But to cement this response, I need to explain *how* we can be reliable in such contexts in arriving at true moral beliefs *given* the truth conditions assigned to moral beliefs by

31 Harman 1977, 4–8.

society-centered theory, and *given* what psychology tells us about processes of belief formation. In other words, I need to show that there is or can be an epistemically relevant connection between our moral beliefs and the moral facts, given the society-centered account of what those facts consist in. I will return to this issue after discussing the remaining three challenges to ethical naturalism.

To begin, I need to discuss the nature of the ‘moral sensitivity’ that, according to the third of the four challenges, is crucially involved in leading us to our beliefs in cases like the cat example. In the cat example, even if the person sees that the children are lighting a cat on fire, she might not take it that what they are doing is wrong. This would be evidence of a lack of moral sensitivity. Ethical naturalists need to explain the nature and epistemic role of moral sensitivity.

It seems to me that there are three aspects to what we have in mind when we speak of moral sensitivity. One is a heightened tendency to notice morally relevant features of a situation. In the cat example, a morally sensitive person will not fail to notice that an animal is being tortured, or that it is screaming or fighting to get away from the children, or that it is terrified. A less sensitive person might not notice these facts about the situation. He might see that the children are lighting the cat on fire without understanding what this will mean for the cat. The second aspect of moral sensitivity is a reliable tendency to draw the correct moral conclusion from the noticed morally relevant features of situations, and to draw this conclusion as promptly as is morally appropriate. In many circumstances, this drawing of the correct conclusion would not involve conscious reasoning. In fact, in some circumstances, a need to reason consciously from morally relevant features of a situation to the moral conclusion would be a sign of moral insensitivity. In the cat example, a morally sensitive person who could see what was happening would immediately realize that it was morally unacceptable. In other circumstances, although some reasoning would be appropriate, protracted reasoning would be untoward. This explains why I say that moral sensitivity involves a tendency to draw the correct conclusion “as promptly as is morally appropriate.” The third aspect of sensitivity is a reliable tendency to be motivated in the morally appropriate way. A morally sensitive person in the cat example would want to stop the children from hurting the cat.

These features can come apart, but in a morally sensitive person they do not come apart. A morally sensitive person in the cat example would notice that a cat is being hurt, would draw the obvious conclusion that the children are doing something wrong, and would want to help the

cat. A less sensitive person might notice that a cat is being hurt without drawing the obvious conclusion that the children are doing something wrong, or this conclusion might dawn on him several minutes later, or he might promptly draw the conclusion but callously walk away.

The epistemic significance of moral sensitivity should be obvious. The first and second aspects are kinds of epistemic sensitivity. They are tendencies to notice morally relevant things and to draw correct conclusions. There are analogous epistemic sensitivities with respect to other subject matters. There are sensitivities to garden variety facts, such as horticultural, geographic, fiscal, and emotional facts. Some people notice flowers and pay attention to their names. They readily recognize flowers when they see them. Some people have a 'geographic sensitivity.' They are keenly aware of where they are and where they are going. Other people are noticeably lacking in geographic sensitivity, getting lost very easily, having difficulty orienting themselves to maps, not knowing which way is north and which way is back the way they came. Children and many adults in our society have a sensitivity to dollar bills, having a keen tendency to notice unattached bills – bills lying on the sidewalk, for example – and to grab them when possible. Some people have a keen awareness of the emotions of other people, being well attuned to the symptoms of how others are feeling and being accurate in the conclusions they draw about their feelings. People who lack this kind of sensitivity to emotions might not notice when a friend is feeling embarrassed or uncomfortable or self-satisfied or content. Moreover, emotional sensitivity is clearly an empirical sensitivity, since the emotions are empirical phenomena. Of course, since the emotions are morally relevant, moral sensitivity will involve emotional sensitivity, but the two kinds of sensitivity are nevertheless distinguishable since one could conceivably be emotionally sensitive without being morally sensitive.

There are motivational aspects of some of the garden variety sensitivities that are analogous to the motivational aspect of moral sensitivity. People with a horticultural sensitivity care about flowers. People with a geographic sensitivity might feel uncomfortable if they do not know where they are or which way is north. People with a keen sensitivity to available dollar bills typically are motivated to grab the bills, and this motivation partly explains their sensitivity. Emotional sensitivity also has a motivational aspect. A person who is sensitive to the emotions of others does not merely detect these emotions with unusual accuracy. She also responds to them in appropriate ways. If someone who is keenly aware of the fear that other people feel uses her awareness to make herself a more

efficient torturer, we would not want to describe her as “emotionally sensitive.” We reserve the term “emotionally sensitive” for people who not only have a keen and accurate awareness of people’s emotions, but who respond to the emotions they perceive with appropriate concern and compassion. This terminological restriction does not change the fact that emotional sensitivity is a sensitivity to an empirical aspect of persons.

It appears, then, that the existence, epistemic significance, and motivational aspects of moral sensitivity are no problem for ethical naturalism. The basic idea is this. We can acquire knowledge or conceptual frameworks, the having of which enables us, in noticing things, to conceptualize them relevantly and to draw correct conclusions about them. We can also come to care about the relevant kinds of things. The example of ‘dollar bill sensitivity’ illustrates this basic idea, and, according to society-centered theory, ‘dollar bill sensitivity’ is relevantly analogous to moral sensitivity. This basic idea will be articulated more fully in what follows.

6. MORAL REASONING

The first two challenges to naturalism turn on the claim that moral reasoning is not a kind of empirical theorizing of the sort that we would expect if naturalism were true. In order to address these objections, I need to begin with a discussion of moral learning.

On any plausible view, we acquire our initial moral attitudes and beliefs early in our lives, through familiar kinds of teaching, training, and experience. We are taught to believe what our parents and other teachers believe, and of course their beliefs are affected by the moral culture of the society. There are constraints on the process that are set by our psychological and other characteristics, and these constraints presumably have an evolutionary explanation. This need not concern us provided that the explanation, whatever it is, does not undermine our justification for believing that the behavioral and judgmental dispositions that we come to have, as a result of the process, can be reliable in leading us to form true moral beliefs.³²

One important factor is that humans have limited psychological and intellectual capacities. We are better able to understand and apply rather simple general maxims than to understand and apply more complex ones, and there is a limit to how many such maxims we can usefully be taught

32 Walter Sinnott-Armstrong pointed out, in correspondence, that evolutionary explanations of processes of belief formation can explain why these processes are sometimes not reliable, as in the case of explanations of visual and cognitive illusions.

to use or trained to follow. Because of these limits, and because also of the variable circumstances we encounter from time to time, we could not be explicitly taught to have all the particular moral beliefs that it will from time to time be appropriate for us to bring to mind. We need either to be taught general rules or to be trained to have certain key judgmental dispositions. We might be taught that torture is wrong, for example, and be taught also to have an aversion to animals being in pain, so that when we see a cat being caused terrible pain we would both have an aversion to this and tend to judge it to be wrong. We would almost certainly be taught simply that torture is wrong rather than being explicitly taught, for each kind of animal that could be tortured, taken one at a time, that it is wrong to torture animals of that kind. That is, we need to be given a relatively small but not insignificant number of rather general behavioral and judgmental dispositions as well as to be brought to believe a number of corresponding general moral maxims or principles. It is reasonable therefore to suppose that the S-ideal moral code would consist of a number of general principles and maxims.³³

The behavioral and judgmental dispositions I have been discussing ground our moral sensitivity, which is our tendency in some circumstances, such as in the cat example, to respond immediately with a moral judgment and an emotional stance. The third challenge to naturalism drew attention to moral sensitivity. The dispositions in question also underwrite our tendency to have moral beliefs even if we are unable to state their truth conditions in helpful terms. The fourth challenge to naturalism doubted that we could have knowledge in such cases.

Let me now turn to the first and second objections, the objections about moral reasoning. The general principles we come to believe license us to draw relevant inferences. The second challenge drew our attention to the immediacy of many such inferences and argued that they are not inferences of the kinds we find in empirical reasoning. We might reason, for instance, that if thus and so were done to a cat, it would experience enormous pain, and so it would suffer torture, and, since torture is wrong, it would be wrong to do thus and so to the cat. We can call this reasoning “subsumption” since it is a matter of subsuming cases under general

33 The S-ideal code obviously would not consist solely of a master standard calling on people to act on the standard or standards, whatever they are, that would be part of the moral code whose currency in the society would best enable the society to meet its needs. If it consisted solely of a master standard, it would be a rather more familiar, substantive, and concrete standard, such as, perhaps, a utilitarian standard.

rules or principles. The first challenge drew our attention to the kind of reflection we engage in when we encounter cases to which our principles do not straightforwardly apply, including cases of special complexity and cases of a kind we have not encountered before. Moral reflection can lead us to extend or to amend our principles or to refine them in the face of anomalies, such as failures of coherence among our principles or between our principles and the nonmoral facts when principles have nonmoral presuppositions. We can reflect on our overall moral view with the goal of increasing its coherence. Reasoning of this kind is given an idealized characterization in Rawlsian wide reflective equilibrium theory, according to which the aim of moral theorizing is to arrive at moral principles that we would accept in a “wide reflective equilibrium.”³⁴ The first of the challenges asserts that moral reflection of this kind is not empirical. It is not an instance of empirical theorizing or reasoning.

The first two challenges to naturalism claim, then, that the characteristic kinds of moral reasoning are not instances of empirical reasoning. They are not instances of ordinary inductive inference, inference to the best explanation, or empirical theorizing. They are rather cases of subsuming a case under a general rule, of seeing a particular as an instance of a kind, or of drawing connections among general principles and refining our principles in order to increase the coherence of our overall view. But these are exactly the kinds of moral reasoning that we should expect, if the S-ideal moral code consists of a plurality of moral principles, as it seems likely that it does. The principles need to be applied to specific cases. Hence subsumptive reasoning is to be expected. Moreover, if the principles are akin to those that make up commonsense morality, they can come into conflict, and they do not always apply in straightforward ways to unfamiliar situations. We might need to reason about unusually complex moral problems, to extend our views to new kinds of cases, and to refine our views when we notice failures of coherence. Hence, reasoning toward equilibrium is to be expected. In short, if society-centered theory is true, we should expect moral reasoning to include at least subsumptive reasoning and reasoning toward equilibrium. It is false that these kinds of reasoning are not to be expected.

Reasoning that is somewhat similar is found in areas that are uncontroversially empirical. For example, in economics we might reason that since recessions occur in thus and so circumstances, and since we are in

34 See Daniels 1979 and Rawls 1971.

circumstances of that kind, we are likely to experience a recession. This is an example of subsumptive reasoning. In this case, the reasoning itself is perhaps a priori, but the premises are empirical. Also in economics, someone might prove a theorem showing that a certain kind of economy would reach an equilibrium in which all markets clear. This reasoning is analogous to moral reasoning toward coherence since it draws connections among various economic principles just as moral reasoning does among moral principles. In this case, again, the reasoning is a priori, but in order to use the theorem to cast light on any actual economy, the premises must be at least approximately true of that economy. And our knowledge that this is so would be empirical. Accordingly, the fact that analogous kinds of reasoning are characteristic of moral reasoning is not an objection to naturalism. The mistake that lies behind the two objections about moral reasoning is to have an overly simple understanding of how we reason about empirical matters.

Of course, as we saw before, society-centered theory suggests that we *can* acquire moral knowledge by acquiring detailed and complex sociological evidence regarding the content of the S-ideal moral code and then inferring various moral propositions from the content of that code. It would not be plausible to hold that *all* moral knowledge is arrived at in this way. Yet if the theory is true, people surely must sometimes deliberate about a moral claim by asking themselves whether a corresponding standard would be part of a moral code that would serve well the needs of society. This point might suggest two additional objections. First, it might seem that if the theory were true, deliberation of this kind would be common. Yet it is not common, and when it does occur, it is not taken to be obviously probative.³⁵ Second, it might seem that if the theory were true, then since we could acquire moral knowledge through complex sociological theorizing, there ought to be a kind of professional moral expertise just as there is, or to the extent that there is, expertise in economics. Yet even if we agreed that moral expertise is possible in principle, we would not expect it to be more common among sociologists than among people in other occupations, and we would not defer to putative moral experts in the way we defer to experts in economics.³⁶

Let me begin with the first objection. It is true that people do not commonly deliberate about whether the content of a moral code would

35 Richmond Campbell urged me to discuss this objection (in personal correspondence).

36 This objection was suggested by comments made by Justin d'Arms and Walter Sinnott-Armstrong, in personal communications.

serve the needs of society. Society-centered theory is not widely accepted, nor is it intuitively obvious. Some people accept theories that conflict with it, and have had their intuitions tutored by their theoretical commitments. But with these qualifications understood, I suggest that deliberation about the content of the S-ideal moral code does not seem intuitively irrelevant when it is couched in ordinary terms rather than the technical terms of the theory. I cite a newspaper article by an English vicar who argues explicitly for “the old moral universals” on the basis of the impact of their currency on the needs of society.³⁷ Anthropologists have argued in similar ways.³⁸ I myself have argued for various moral views, including views about the treatment of animals, abortion, civil liberties, and the legitimacy of the state, from the basic society-centered theory.³⁹ Arguments of this kind admittedly are rather rare. But society-centered theory suggests why they are rare and suggests that they will be rare even in ideal circumstances. For, in ideal circumstances, the S-ideal moral code would have currency in society S, and this means that people in S would generally have internalized its standards and would tend to reason from them in deciding what to do rather than to reason directly from society-centered theory. Even in less than ideal circumstances, as I have been arguing, we are taught rather general moral rules and tend to reason from them to the extent that we have internalized them.

As for the second objection, I want to insist, to begin with, that the idea that there can be moral expertise is not alien and should not be surprising. Millions of Christians and Muslims in the world believe there is moral expertise and defer to the views of the people they take to be experts. Nor is the idea of moral expertise necessarily based in religious views. Indeed, the idea that there can be moral expertise should be no more surprising than the idea that there can be moral obtuseness. Some people are morally vicious and insensitive, and it is possible, at least in principle, that there are people who are especially virtuous and who are unusually sensitive to morally significant considerations so that they are better judges of right and wrong than the rest of us. It seems to me that it would be difficult to deny this. Indeed, it seems to me that the denial that there can be moral expertise would have to be grounded in an antirealist metaethical view

37 Mullen 1983.

38 Aberle et al. 1950. They discuss sexual abstinence in relation to the needs of society (103–104).

39 See the discussion of moral issues in Copp 1995, 201–209, 213–216. See also Copp 1999.

such as, perhaps, J. L. Mackie's error theory.⁴⁰ Not even noncognitivist or expressivist theories such as the theories of Simon Blackburn and Allan Gibbard are committed to denying the possibility of moral expertise.⁴¹ And typical normative theories, such as utilitarian and Kantian theories, imply that it is possible in principle for some people to be better judges of right and wrong than the rest of us. It is true that we are morally responsible for our own decisions. We cannot avoid this responsibility by deferring to a putative expert. If we decide to accept someone's advice, we can be held responsible both for our choice of advisor and for acting on the advisor's advice. If we act on bad advice, we can be blameworthy for doing so, although if we acted in good faith and were careful in choosing an advisor, the blame might be mitigated. In any event, the important point is that it should not be surprising that my position implies there can be moral expertise, nor is this an objection to it.

It is true of course that a sociologist is no more likely to exhibit moral expertise than is anyone else. This is because the moral education and experience of sociologists, and their moral views, are of the same kind as the moral education and experience, and the moral views, of people in other occupations. According to society-centered theory, sociological evidence can be especially relevant to assessing the content of the S-ideal moral code, but the evidence has not been assembled, and, moreover, sociologists are not more likely to accept the society-centered theory than are other people.

7. HOW ARE MORAL BELIEFS LINKED TO THE MORAL FACTS?

Earlier, I set aside questions about the existence of an epistemically relevant connection between moral beliefs and the moral facts and I also set aside questions about the epistemic credentials of the psychological processes by which we come to have moral beliefs. These questions now need to be answered. The problem is to explain how our capacity for moral sensitivity, as well as our capacity to reason about moral facts, come to be linked to what are *actually* moral facts rather than merely to the facts taken in a given society to be morally relevant. The problem is not merely academic, for there have been unjust societies in which corrupt moral beliefs were widespread. The underlying issue is to explain how our moral

40 Mackie 1977.

41 Blackburn 1988 and 1993; Gibbard 1990.

beliefs come to be linked epistemically to what are actually moral facts in a way that underwrites the possibility of moral knowledge. This is the central problem for moral epistemology.

We accept a number of moral principles. Our acceptance of them is a result of the initial moral teaching we were given combined with subsequent experience and reasoning. The principles we accept might or might not be true. And even if some principle we accept is not true, it might be an approximation to the truth. For example, we might come to think that torture is wrong except for the torture of nonhuman animals. I take it that this is at least an approximation to the truth, even though there is no justifiable exception that permits the torture of nonhuman animals, for, I assume, torture *is* wrong.⁴² Suppose then that the moral principles we believe as a result of this process of initial teaching and subsequent experience and reasoning are true or approximately true, and suppose that the judgmental dispositions we have as a result of these processes tend to lead us to make true judgments. If these suppositions are true, then, I believe, the true moral beliefs we may come to have as a result of reasoning from the principles we accept, or as a result of appropriate exercises of our moral sensibility, might count as knowledge. The explanation for the truth of the suppositions must be of the right kind, however. That is, the truth or approximate truth of the moral principles we accept, and the accuracy of our judgmental dispositions, must be due at least in part to the existence of an appropriate kind of causal connection between the moral facts and the truth or approximate truth of the relevant beliefs. Can we make it plausible that there might be such a connection?

The chief problem here for society-centered theory is to explain how it could be that the truth or near truth of our moral perspective might be due in part to the fact that the corresponding moral code is, or approximates to, the S-ideal moral code. To begin with, it seems unlikely that the truth or near truth of a moral perspective would be entirely accidental. If the principles we believe are true or approximately true, and if a large proportion of our basic moral beliefs is true as a result, this is *not* likely to be purely coincidental. It is likely to be a result, at least in part, of the fact that the truth conditions of the beliefs do in fact obtain. Indeed, on the assumption that society-centered theory is correct, it is likely to be a result, at least in part, of the fact that the standards corresponding to our beliefs would belong to, or be approximations of standards that would

42 For an argument that society-centered theory supports a prohibition on cruelty to animals, see Copp 1995, 204–207.

belong to, the S-ideal moral code. But even if all of this is plausible, a naturalist still needs to explain the nature of the mechanism or mechanisms, the operation of which could bring it about that our moral perspective approximates to the S-ideal moral code of our society.

There is, first, I suspect, a comparative mechanism. If society-centered theory is correct, societies do better at coping with their problems, other things being equal, to the extent that their members have true moral beliefs and subscribe to corresponding moral standards. And over time, other things being equal, we can expect the societies that are more successful at meeting their needs to thrive. Societies in which a large proportion of people's moral beliefs are true or approximately true and in which people subscribe to corresponding standards should do better at meeting their needs in their ecological circumstances than otherwise would be the case, other things being equal. Such societies should tend to absorb the populations of societies in the same or similar ecological circumstances in which the moral culture is less accurate to the underlying S-ideal moral code. Or the moral cultures of such societies should tend to be exported to societies in similar circumstances that are doing less well at meeting their needs because of the content of their societal codes. Moreover, many of the standards that would usefully be subscribed to are obviously so, and subscription to them by the members of many smaller groups would also contribute to the success of these other groups. For these reasons it is likely that, with time, in at least some societies, and other things being equal, the prevailing moral outlook will tend toward the S-ideal code, the currency of which would best enable the society to meet its needs.

I believe there is also likely to be a feedback mechanism, a mechanism that involves a feedback between the dominant moral perspective or perspectives of a culture and the corresponding society's ability to meet its needs. In favorable circumstances – but, unfortunately, not in all circumstances – the mechanism would tend to bring about changes in the direction of the S-ideal moral code in cases in which the society's ability to meet its needs could be improved by such changes. And it would tend to stabilize the dominant moral perspective in cases in which the corresponding code approximated sufficiently to the S-ideal code.⁴³

I take it that we have observed moral progress in the dominant moral perspective in American society over the past century. We have seen

43 For a more thorough discussion of a similar proposed feedback mechanism, see Railton 1986, 192–197, 204–207.

improvements in people's attitudes in race relations, in gender relations, in reproductive matters, in treatment of youth, and so on. And I would argue that these changes are improvements when judged by the criterion proposed by society-centered theory. The pressing question is whether these changes were due at least in part to the effect on the moral culture of the fact that the society was better able to meet its needs after the changes than before. This is a large issue that goes beyond the scope of this chapter. But I think there is evidence that this was so. Race relations and gender relations improved at least in part, I believe, because the society could ill afford to waste the talent of African Americans and women who were often shunted into demeaning or undemanding jobs or unpaid work. In order to draw from these talented people the benefits that their full participation could offer to the society, social attitudes needed to change so that full participation would be accepted. I think that many people understood these facts. Of course, they perhaps would have described themselves, and not inaccurately, as seeing the injustice of discrimination. And, partly for this reason, efforts were made to change people's attitudes, partly through legal reform and partly through more diffuse social channels. The exact mechanisms are perhaps poorly understood, and it is certainly true that there is a level of description on which the mechanisms have little to do with the needs of society. I believe, however, that there is also a level of description on which the mechanisms are affected by the needs of society, as I have attempted to illustrate.

The argument I have been giving depends on two suppositions. First is the supposition that the moral principles we accept as a result of the processes of moral teaching, reasoning, and experience are true or approximately true, and second is the supposition that the judgmental dispositions we have as a result of these processes tend to lead us to make true judgments. I argued that, if these suppositions are true, the explanation for this might well be of the right kind. It might well be that the comparative and feedback processes I described have resulted over time in the truth or approximate truth of our moral perspective. And, if this is so, then, I claim, true moral beliefs we have as a result of moral reasoning, or as a result of exercises of our moral sensibility, might count as knowledge. This is not to say, of course, that they *would* count as knowledge. It might be, for instance, that we believe truly that capital punishment is wrong, and the explanation for this might be of the right kind, but if we ignore the contrary beliefs of reasonable people who disagree with us, and if we have no response to their objections, then we should not be said to have

knowledge that capital punishment is wrong.⁴⁴ My claim is simply that, under favorable conditions, we *can* have moral knowledge.

8. NATURALIZED EPISTEMOLOGY REVISITED

We have now addressed the four challenges to naturalism. And I have argued that ethical naturalism has nothing to fear from naturalized epistemology. The main lesson of naturalized epistemology, I said, is that a semantical and metaphysical theory of a given subject matter can be plausible only if it meshes suitably with a psychologically and philosophically plausible epistemology. It would indeed be difficult to deny this.

There are philosophers who deny it, in effect, since they postulate a special faculty of ‘intuition’ by which we acquire moral knowledge. They might postulate such a faculty, despite the lack of any psychological evidence that one exists, on the basis of arguments to the effect that such a faculty is required if we assume that there is moral knowledge. Moves of this kind run afoul of the scientific prioritism that is implicit in naturalized epistemology. I conceded that I know of no arguments for prioritism that would convince someone inclined to reject it, but naturalists will want to accept it. The bare idea that some of our moral beliefs are justified noninferentially does not run afoul of scientific prioritism, however, and this idea might be viewed as a weak form of intuitionism. Since I think that true beliefs we come to have as a result of appropriate exercises of our moral sensibility can count as noninferential knowledge, my arguments appear to support a minimal form of naturalistic ‘intuitionism.’⁴⁵

Naturalized epistemology does imply that the acceptability of my responses to the four epistemological challenges depends on their psychological plausibility. Of course, their psychological plausibility is an empirical matter that is beyond the scope of my work, but the kind of psychological considerations that would undermine my responses should be clear. I shall discuss one example, a theory in cognitive science that might appear to conflict with society-centered theory.

Society-centered theory makes prominent use of the ideas of a moral standard and a moral code, where a moral standard is a kind of rule,

44 For discussion of such cases, see Copp 1996.

45 Robert Audi describes a “modified” form of ethical intuitionism (1996). Given the restrictions he imposes on the idea of an “intuition,” I do not know whether he would describe my view as intuitionistic.

and a moral code is a system of such rules. I argued, moreover, that the processes of moral learning result in our accepting a number of general maxims or principles, which might also be called “rules,” and I described moral reasoning as involving, among other things, subsuming cases under such principles. These features of my view might appear to put it in conflict with accounts of cognitive processing that Paul Churchland and Alvin Goldman think are supported by research in cognitive science and artificial intelligence. On the basis of “prototype” theory, Churchland has suggested that moral concepts should be construed as “prototypes” rather than as packages of necessary and sufficient conditions. Echoing particularists, he says that “One’s ability to recognize instances of cruelty, patience, meanness, and courage, for instance, far outstrips one’s capacity for verbal definition of these notions.” Hence, he suggests, “it is just not possible to capture, in a set of explicit imperative sentences or rules, more than a small part of the practical wisdom possessed by a mature moral individual.” He concludes that “a rule-based account of our moral capacity” is mistaken and should be replaced with an account that exploits the idea of a “hierarchy of learned prototypes.”⁴⁶ Alvin Goldman describes a view he calls the “exemplar view” in a similar way as holding that “concepts are (sometimes) represented by one or more of their specific exemplars, or instances, that the cognizer has encountered.” On the basis of the exemplar view, he argues that moral learning might consist primarily in “the acquisition of pertinent exemplars or examples” rather than the learning of rules.⁴⁷ It appears, then, that if Churchland and Goldman are correct, exemplar or prototype theory threatens my account of moral learning and reasoning, and might even threaten society-centered theory itself.

Let me set aside questions about the theoretical plausibility of prototype or exemplar theory, and about the evidential support they enjoy. This is not settled science. The important point, however, is that we do not have to choose between exemplar or prototype theory and the idea that we accept moral rules any more than we have to choose between these theories and the idea that we have concepts. The theories provide accounts of what concepts are, or of how they are represented. They do not imply that we have no concepts. Similarly, we should view the theories as offering accounts of how moral rules are represented, or what their acceptance consists in, rather than as showing that we do not accept any moral rules.

46 Churchland 1996, 101, 102, 106. For particularism, see, e.g., Dancy 1993.

47 Goldman 1993, 340–341. Goldman cites Estes 1986; Stich 1993; and Medin and Schaffer 1978.

Perhaps, for example, a person who accepts a 'rule' that prohibits cursing has a concept of cursing that either is, or is represented by, a prototype or exemplar, and perhaps the rule is best construed as prohibiting actions that are relevantly similar to the prototype. On this understanding, the currency of moral code would depend on our having sufficiently similar prototypes. For example, the currency of a moral code calling for patience with children and precluding cruelty to animals might depend on our having sufficiently similar prototypes of patience and cruelty. But none of this is an argument against the idea that we accept moral rules and none of it constitutes an objection to anything I have said.

In any event, I have argued that ethical naturalism has nothing to fear from naturalized epistemology. I have argued that the moral epistemology I relied on, in responding to the four epistemological challenges to naturalism, is relevantly similar to the epistemology of various other subject matters. Let me conclude by spelling out some of the details.

9. CONCLUSION

As we saw, society-centered theory holds that true basic moral claims are true in virtue of the obtaining of relevant facts about the content of the S-ideal moral code. But our moral beliefs are not typically grounded in evidence regarding the content of the S-ideal moral code. Few people accept society-centered theory, and few of us are in a position to decide what to believe morally on the basis of evidence regarding which moral standards would be part of the S-ideal moral code. Our early moral training and subsequent experience and reasoning do not typically give us any direct evidence about the content of the S-ideal moral code. Yet, if I am correct, we can have moral knowledge even if the relevant beliefs are not grounded in evidence that relevant corresponding moral standards would be part of the S-ideal moral code, and even if no one in the causal ancestry of the relevant beliefs has ever had such evidence. In this respect, moral knowledge is similar to empirical knowledge in many other areas of thought. You can know that dollar bills are green even if you do not believe this on the basis of articulated evidence that its scientifically explicit truth conditions obtain, and even if no one in the history that led to your believing it believed it on such a basis. All of this is compatible with society-centered theory.

Whether a given moral belief qualifies as knowledge depends on whether there is an epistemically relevant link of the right kind between the belief and the facts in virtue of which it is true. In cases where we

have moral knowledge, the epistemically relevant link between our moral belief and the natural fact in virtue of which it is true is brought about through the genesis of the belief, which traces to our early moral learning and to subsequent experience and reasoning. If the moral perspective we have acquired through these mechanisms is true or approximately true, and if this is nonaccidental and due at least in part to the fact that the corresponding moral code is, or approximates to, the S-ideal moral code, then a true moral belief that we have at least partly as a result of having this moral perspective might qualify as knowledge.

I argued before that the S-ideal code likely would include a variety of general moral principles rather than a single master principle. So, if our outlook approximates to the S-ideal code, we accept a variety of principles. In some circumstances, these principles might point us in different directions. When they do, we need to look for a single prescription, and we can do so by reasoning in the familiar way that begins with principles and prescriptions that we accept and looks for a prescription for the case at hand that best coheres with them. What is to be expected is that we have certain dispositions to judge, respond, and reflect, the nature of which is due to the combined effect of our initial moral teaching and training and subsequent experience and reasoning. Our dispositions to judge and respond enable us to have a special moral sensitivity to the morally relevant characteristics of situations. Together with dispositions to reflect, they explain how moral perception is possible, how immediate inferences are grounded in general moral beliefs or dispositions to believe, and why reasoning that tends toward wide reflective equilibrium is characteristic of moral reasoning.

Suppose then that your belief that capital punishment is wrong is in fact true. You believe this about capital punishment, let us say, on the basis of reasoning from moral premises your acceptance of which traces back ultimately to a moral outlook that you were taught in childhood. Suppose that this reasoning has in fact made your overall outlook more coherent than it would otherwise have been and has also made it a closer approximation to the S-ideal moral code, and suppose that the moral outlook you were taught in childhood also approximated to the S-ideal code and did so partly as a result of the operation of feedback and comparative mechanisms of the sort that I have described. Under these circumstances, I think your belief could qualify as knowledge. If it did, it would qualify as empirical knowledge since it would rest on your experience and on the experience and observations of others and on the existence of a

mechanism that tends to bring our moral perspective into line with the S-ideal moral code.

From the first-person perspective, matters seem different from how they are seen from the perspective of metaethical theory. It seems that we can sometimes just *see* what would be right or wrong, without being aware of drawing any inferences at all. From the first-person perspective as well it seems that we can draw immediate inferences that are not like the inferences we make in empirical reasoning. Moreover, it seems that we can be morally perplexed even when all the facts are in. And in such cases of perplexity, it seems that our reasoning about what to do is rather more like a priori reasoning than it is like the reasoning we should expect to be engaged in if the issues were fundamentally empirical. These phenomena can all seem to undermine naturalism, yet I have argued that naturalism can make sense of them. Indeed, among cognitivist positions, it seems that ethical naturalism is the only kind of view that can make sense of all of this within the strictures of naturalized epistemology.

REFERENCES

- Aberle, D. E., A. K. Cohen, A. K. Davis, M. J. Levy, Jr., and F. X. Sutton. 1950. "The Functional Prerequisites of a Society." *Ethics* 60: 100–111.
- Audi, Robert. 1996. "Intuitionism, Pluralism, and the Foundations of Ethics." In Sinnott-Armstrong and Timmons 1996: 101–136.
- Benacerraf, Paul. 1973. "Mathematical Truth." *Journal of Philosophy* 70: 661–679.
- Blackburn, Simon. 1988. "How to Be an Ethical Antirealist." *Midwest Studies in Philosophy* 12: 361–375.
- . 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- Churchland, Paul. 1996. "Neural Representation and the Social World." In *Mind and Morals: Essays on Cognitive Science and Ethics*, ed. Larry May, Marilyn Friedman, and Andy Clark, 91–108. Cambridge: Bradford Books of MIT Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In Sinnott-Armstrong and Timmons 1996: 243–266.
- . 1997. "Does Moral Theory Need the Concept of Society?" *Analyse und Kritik* 19: 189–212.
- . 1998. "Morality and Society – The True and the Nasty: Reply to Leist." *Analyse und Kritik* 20: 30–45.
- . 1999. "The Idea of a Legitimate State." *Philosophy and Public Affairs* 28: 3–45.
- Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Basil Blackwell.
- Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *Journal of Philosophy* 76: 256–282.

- Estes, William. 1986. "Array Models for Category Learning." *Cognitive Psychology* 18: 500–549.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Goldman, Alvin I. 1979. "What Is Justified Belief?" In *Justification and Knowledge*, ed. George S. Pappas, 1–23. Dordrecht, Holland: D. Reidel.
- . 1993. "Ethics and Cognitive Science." *Ethics* 103: 337–360.
- Harman, Gilbert. 1977. *The Nature of Morality*. New York: Oxford University Press.
- Kim, Jaegwon. 1993. "What Is 'Naturalized Epistemology'?" In *Supervenience and Mind: Selected Philosophical Essays*, 216–236. Cambridge: Cambridge University Press.
- King, Jeffrey C. 1994. "Can Propositions Be Naturalistically Acceptable?" *Midwest Studies in Philosophy* 19: 53–75.
- Kornblith, Hilary. 1985a. "Introduction: What is Naturalistic Epistemology?" In Kornblith 1985b: 1–14.
- Kornblith, Hilary, ed. 1985b. *Naturalizing Epistemology*. Cambridge: MIT Press.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Medin, Douglas, and M. M. Schaffer. 1978. "A Context Theory of Classification Learning." *Psychological Review* 85: 207–238.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt. Cambridge: Cambridge University Press, 1993.
- Mullen, Peter. 1983. "What's Wrong Can Never Be Right." *Manchester Guardian Weekly* (December 4): 4.
- Quine, W. V. O. 1969a. "Epistemology Naturalized." In Quine 1969c: 69–90.
- . 1969b. "Natural Kinds." In Quine 1969c: 26–68.
- . 1969c. *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- . 1990. *In Pursuit of Truth*. Cambridge, Mass.: Harvard University Press.
- Railton, Peter. 1986. "Moral Realism." *Philosophical Review* 95: 163–207.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Sinnott-Armstrong, Walter, and Mark Timmons, eds. 1996. *Moral Knowledge? New Readings in Moral Epistemology*. New York: Oxford University Press.
- Steiner, Mark. 1973. "Platonism and the Causal Theory of Knowledge." *Journal of Philosophy* 70: 57–66.
- Stitch, Stephen. 1993. "Moral Philosophy and Mental Representation." In *The Origin of Values*, ed. Michael Hechter, Lynn Nadel, and Richard E. Michod, 215–228. Hawthorne, N.Y.: Aldine de Gruyter.
- Stroud, Barry. 1985. "The Significance of Naturalized Epistemology." In Kornblith 1985b: 71–89.

Moral Naturalism and Self-Evident Moral Truths

1. THE ISSUE

It is intuitively plausible that there are substantive moral propositions that are ‘self-evident.’ It is plausible, for example, that, “other things equal, it is wrong to take pleasure in another’s pain, to taunt and threaten the vulnerable, to prosecute and punish those known to be innocent, . . . to sell another’s secrets solely for personal gain,” and “to torture others just for fun.”¹ It is plausible that these propositions are true, and it is plausible that they are self-evident. In what follows, I refer to them as “the common sense principles.” And I will call the thesis that some such propositions are self-evident “the self-evidence thesis.”

It is not entirely clear how to understand the idea of a self-evident proposition. Intuitively, a self-evident proposition is one that is obviously true without the need for any proof or argument. But the term “self-evident” is used as a technical term in philosophy, and philosophers have meant different things by it. Russ Shafer-Landau, who gives the common sense principles as examples, proposes a stipulative definition.² Expressed

A version of this chapter was presented to the March 2006 Meetings of the Pacific Division of the American Philosophical Association in Portland, Oregon. I am grateful to Mark Timmons for organizing the session and to Russ Shafer-Landau, Sarah Stroud, and the members of the audience for very helpful discussion. I also discussed the chapter with a group of colleagues at the University of Florida, and I am grateful for their helpful comments and acute criticisms. I would especially like to thank John Biro, Stuart Duncan, Tom Hurka, Marina Oshana, Greg Ray, Russ Shafer-Landau, Jon Tresan, and Gene Witmer. I am also grateful to Walter Sinnott-Armstrong for very useful comments.

1 Shafer-Landau 2003, 248.

2 Ibid., 247.

informally, his idea is that “once one really understands” the common sense principles, “(including the *ceteris paribus* clause),” one is justified in believing them.³ Robert Audi proposes a somewhat different definition. He suggests that a self-evident proposition is such that anyone who “adequately understands” it would be justified in believing it and would know it if he believed it on the basis of this understanding.⁴

On either of these definitions, it may seem obvious that if there are any self-evident moral principles, they are knowable a priori, or they are at least a priori justifiably believed. For if merely adequately to understand a proposition is sufficient to justify believing it, then no empirical evidence is required to justify believing it, once one understands it adequately. And it would seem to follow that one can be justified a priori in believing it. This brings us to an important issue for moral naturalism, the issue that will be the central focus of this chapter.

The defining thesis of moral naturalism is the thesis that the moral properties are ‘natural properties.’ It is not clear how best to understand the idea of a natural property, but, plausibly, it is the idea of a property that has the same metaphysical and epistemological status as the ordinary unexceptional properties of ordinary objects in the world around us. It is plausible that any significant, substantive knowledge or justified belief we have about the instantiation of such properties is empirical or a posteriori. That is, for any *synthetic* proposition *p* about the instantiation of a such a property, it is plausible that if we have knowledge or justified belief that *p*, our knowledge or justification is a posteriori. But then since moral naturalism views moral properties as having the same metaphysical and epistemological status as ordinary unexceptional properties, moral naturalism seems committed to viewing all of our knowledge or justified belief of synthetic moral propositions as a posteriori.

If this is correct, it appears that moral naturalism is incompatible with the self-evidence thesis. The self-evidence thesis says that some substantive moral principles are self-evident, which, on the above argument, entails that they are a priori knowable or justifiably believed. But, given the above reasoning, moral naturalism implies that all of our knowledge or justified belief of synthetic moral propositions is a posteriori. On this showing, then, a naturalist cannot accept the self-evidence thesis unless she is prepared to argue that the only moral propositions that are

3 Ibid., 248.

4 Audi 2004, 48–49.

self-evident are conceptual truths. Shafer-Landau has argued in roughly this way from the self-evidence thesis, together with the implicit claim that there are synthetic moral truths that are self-evident, to the untenability of moral naturalism.⁵ I contend that the argument is unsuccessful. In order to evaluate it, however, we need to explain what moral naturalism amounts to as well as to explicate the idea of a self-evident proposition. The argument may be more difficult to resist on the epistemic conception of a natural property that I have defended elsewhere⁶ than it would be on a purely metaphysical conception of a natural property. But I shall argue that even on my conception of a natural property, and even on Audi's or Shafer-Landau's account of self-evidence, and even if the common sense principles are synthetic, a naturalist can agree that these principles are self-evident. The self-evidence thesis is compatible with moral naturalism.

In the [next section](#), section 2, I briefly explain the epistemic conception of moral naturalism. In section 3, I explore Audi's and Shafer-Landau's accounts of self-evidence and propose a revised account based on their views. In section 4, I concede that the common sense principles are synthetic. In section 5, I discuss whether, if a proposition is self-evident, it follows that it is a priori knowable or justifiably believed. In section 6, I argue that even if the common sense principles are synthetic, the claim that they are self-evident is compatible with naturalism. In section 7, I explain how a specific version of moral naturalism can make sense of this. I conclude with a brief summary.

5 Shafer-Landau argues for an epistemic construal of moral naturalism according to which moral properties are natural properties in that "the fundamental truths that describe [their] nature, and the conditions under which [they are] instantiated, are discoverable in an exclusively a posteriori way" (2003, 61). He then says that his own view, which he says he defends in chapter 11, "is that we can discover fundamental ethical truths in an a priori fashion" (2003, 61, see 65). The thesis he explicitly argues for in chapter 11, however, is the self-evidence thesis. So he appears to be assuming that if there are self-evident moral propositions, they are a priori knowable. Moreover, if he had in mind that these were conceptual truths, he surely would have said this. So I take his view to be that the self-evidence thesis entails that there is a priori moral knowledge, which, given that the propositions in question are synthetic, entails the untenability of moral naturalism. (Shafer-Landau conceded, in discussion (March 2006), that he was assuming in the book that self-evident propositions are a priori knowable.) Audi appears to hold that all propositions that are self-evident in his sense are a priori knowable. And he claims that some synthetic propositions are self-evident, so that his account is able to vindicate the idea of the synthetic a priori. Audi 1997, 62 n. 24, and 102; Audi 2004, 211 n. 19.

6 See chapter 1.

Moral naturalism is a kind of moral realism. A moral realist holds, among other things, that there are moral properties, such as goodness and rightness. The moral naturalist adds that these properties are *natural* properties. Of course the naturalist needs to explain the idea of a ‘natural’ property. I am going to work with the epistemic conception of a natural property that I have proposed elsewhere.⁷ The tension between moral naturalism and the self-evidence thesis seems especially acute if we have in mind an epistemic conception of natural properties.

Shafer-Landau also proposes an epistemic conception. He suggests that natural properties are “those that would figure ineliminably in perfected versions of the natural and social sciences.” He argues that the crucial feature of the sciences is the “exclusion of a priori knowledge of fundamental scientific truths.” Accordingly, he suggests, “the essence of a natural property is that the fundamental truths that describe its nature, and the conditions under which it is instantiated, are discoverable in an exclusively a posteriori way.”⁸

I want to quibble about two points. First, it seems to me that a naturalist need not hold that ethical properties would “figure ineliminably” in perfected versions of the sciences. There are ordinary properties of artifacts and the like, such as the property of being a two by four, the property of being a stick, and the property of being two meters long. These properties are natural properties, we should say, and we should not rest this claim on the idea that they will “figure ineliminably” in a perfected version of a science. Second, it would be a mistake for a naturalist to insist that all moral knowledge is a posteriori. A naturalist can acknowledge that conceptual truths with moral content are knowable a priori. If it is a conceptual truth that murder is wrong, we can know a priori that murder is wrong.⁹ This does not mean that wrongness is not a natural property. Similarly, we can know a priori that vixens are foxes, but this does not mean that the property of being a fox is not a natural property.

I propose the following, as an initial characterization: A natural property is a property such that any true *synthetic* proposition about its

7 Ibid.

8 Shafer-Landau 2003, 59, 61.

9 For my purposes it does not matter whether this is correct. If it is not correct, then let me introduce the concept of a schmurder by saying that a schmurder is a wrongful killing. It is, then, a conceptual truth that schmurder is wrong. Artificial examples aside, one may doubt whether there are any conceptual truths, but the important point is that naturalism is compatible with recognizing the existence of conceptual truths.

instantiation is only knowable a posteriori. On this proposal, the disagreement between naturalism and nonnaturalism is best understood as a disagreement about the existence of synthetic a priori moral knowledge. That is, on this proposal, moral naturalism is best taken to imply that all knowledge of synthetic moral truths is a posteriori.¹⁰

For present purposes, I stipulate that a proposition is *synthetic* just in case it is neither a conceptual truth nor the negation of a conceptual truth. Moreover, ‘theorems’ as well as ‘axioms’ qualify as conceptual truths for these purposes. That is, if a set of conceptual truths is such that it is a conceptual truth that the propositions in the set entail a further proposition, then we should count the latter as a conceptual truth. Call it a “derivative conceptual truth,” and call conceptual truths that are not derivative in this way “strict conceptual truths.”

My initial characterization of the idea of a natural property needs to be modified in light of a distinction that has been drawn by Hartry Field. Field says,

Let’s define a *weakly a priori* proposition as one that can be reasonably believed without empirical evidence; an *empirically infeasible* proposition as one that admits no empirical evidence against it; and an *a priori* proposition as one that is both weakly a priori and empirically infeasible.¹¹

I will call Field’s third kind of proposition “strongly a priori.”

10 For a more detailed discussion, see chapter 1. We need to distinguish a proposal about how best to explicate moral naturalism and its idea of a natural property – an explicating proposal – from a proposal about how to classify philosophers’ positions as naturalist or nonnaturalist – a classification proposal. According to proposals of both kinds, moral naturalism combines moral realism with the doctrine that moral properties are natural. An explicating proposal adds that a property is natural just in case it is N, for some appropriate N. Such a proposal will not be plausible unless the thesis that moral properties are N is entailed by, or is at least compatible with, most positions that would standardly be classified as versions of moral naturalism. Despite this, however, such a proposal does not commit us to the corresponding classification proposal that a philosopher’s theory is a version of moral naturalism if and only if it entails that moral properties are N. This is because a philosopher might reject our explicating proposal even if she holds that moral properties are natural properties and even if she classifies her own view as naturalistic. It might be unclear how best to classify the position of such a philosopher, but we should not deny that she is a naturalist merely because she denies that moral properties are N. The thesis that moral properties are N is, by our lights, the thesis that best explicates the idea that moral properties are natural, but the fundamental idea that all moral naturalists share is that moral properties are natural, not that moral properties are N. A classification proposal should take into account how a philosopher would characterize her own position as well as her other views about the nature of moral properties. I am grateful to Greg Ray and Gene Witmer for helpful discussion of this issue.

11 Field 2000, 117. Field emphasizes that by “reasonable” he means “epistemically reasonable” (117 n. 2).

It seems to me that a naturalist can acknowledge that certain synthetic moral propositions may have a kind of default credibility in virtue of which they qualify as weakly a priori.¹² Consider, for example, the proposition that friendship is good. It is synthetic, yet it plausibly has a kind of default credibility for people sharing our culture, given the significance that our culture accords to friendship. This tells us nothing about the nature of the property of goodness, so it seems to be compatible with viewing goodness as a natural property. Yet if one denied that the proposition that friendship is good is empirically defeasible, while acknowledging that it is synthetic, it seems to me that one would have to think that goodness is different in nature from the ordinary properties of friendship, such as, say, the property of being pleasing, synthetic propositions about which are empirically defeasible. It seems to me, therefore, that if goodness is a ‘natural’ property, synthetic propositions about its instantiation are empirically defeasible.

For this reason, I need to amend my initial characterization. I propose the following: A natural property is a property such that any synthetic proposition about its instantiation is empirically defeasible.¹³ We should understand moral naturalism accordingly as implying that any synthetic proposition about the instantiation of a moral property is empirically defeasible; no such proposition is strongly a priori.¹⁴

Let me clarify the notion of empirical indefeasibility. It is arguable that any synthetic proposition is empirically defeasible because evidence that it is widely believed to be false would be empirical evidence against it.¹⁵ A nonnaturalist might object that the fact that there can be evidence of this kind against the truth of a synthetic moral proposition has no bearing whatsoever on the issues that divide naturalism from nonnaturalism since it is completely independent of the truth conditions of the

12 I have argued this in chapter 1.

13 Saul Kripke argues that there are certain propositions that are contingent a priori, given the way the reference of certain terms has been fixed. See Kripke 1980, 54–57, esp. 56. For our purposes, it is important whether the propositions in question are synthetic. If they are synthetic, then on my proposal, it is important whether they are empirically defeasible. One example Kripke uses is the proposition that S is one meter long, where S is the standard meter bar that has been used by someone to fix the reference of “meter.” Kripke argues that this person can reasonably believe the proposition “without further investigation” (56). Even if so, and even if the proposition is synthetic, I do not believe it is strongly a priori. I will return to this example in section 5. I am grateful to Michael Jubien, Greg Ray, Michael Ridge, and Jon Tresan for helpful discussion.

14 Jonathan Dancy argues that all substantive moral knowledge is contingent a priori (2004, 146–148). I owe this reference to Michael Ridge. I maintain that if the propositions Dancy has in mind are synthetic, a naturalist should deny that they are strongly a priori.

15 See chapter 1.

proposition.¹⁶ Let me therefore draw a distinction. I will say that the fact that people believe *p* to be false is merely “indirect” evidence against *p*. Its status as evidence against *p* would be undermined or defeated by evidence that the people who believe *p* to be false are epistemically unreasonable to do so. That is, its status as evidence against *p* depends on the epistemic credentials of these people’s belief that not *p*. Let us then stipulate that “direct” empirical evidence against *p* is empirical evidence that *p* is false that is not indirect in this sense, and let us stipulate that a proposition counts as “empirically indefeasible” just in case there can be no *direct* empirical evidence that it is false.¹⁷ A strongly a priori proposition is both weakly a priori and empirically indefeasible in this sense.

3. THE IDEA OF A SELF-EVIDENT PROPOSITION

To the best of my knowledge, the most thoroughgoing investigation of the idea of a self-evident proposition is found in the work of Robert Audi. He proposes the following definition:¹⁸

A self-evident proposition is (roughly) a truth such that [adequately] understanding it [perhaps following careful reflection] will meet two conditions: that understanding is (1) sufficient for one’s being justified in believing it . . . and (2) sufficient for knowing the proposition provided one believes it on the *basis* of understanding it.

Unfortunately, Shafer-Landau proposes a somewhat different definition. He stipulates that “A proposition *p* is self-evident = *df.* *p* is such that adequately understanding and attentively considering just *p* is sufficient to justify believing that *p*.”¹⁹

16 I am grateful to Tom Hurka and Russ Shafer-Landau for pressing this objection in discussion at the APA session at which this chapter was presented.

17 There is a distinction between evidence that a belief that *p* is unjustified and evidence that the proposition *p* is false. The distinction in the text is between two kinds of evidence that a proposition is false. I am proposing to stipulate that an empirically indefeasible proposition is such that there can be no direct empirical evidence that it is false. Audi claims that certain “beliefs of empirical propositions” are “indefeasibly justified.” The cases he has in mind are first-person beliefs, such as “my beliefs that I have a belief and (arguably) that I exist” (Audi 2004, 213 n. 31). In each of these examples, the *proposition* that is the object of the belief is empirically defeasible. The propositions in question are, respectively, the proposition that Audi has a belief and the proposition that Audi exists, both of which are empirically defeasible.

18 Audi 1997, 45 and 61 n. 24. Cf. Audi 2004, 48–49. See also Audi 1996, 114.

19 Shafer-Landau 2003, 247.

To simplify my discussion, I need to work with a single definition. I propose the following, which retains the most attractive features of each of Audi's and Shafer-Landau's definitions:

A proposition *p* is self-evident just in case *p* is a truth such that adequately understanding and attentively considering just *p* is sufficient for one to be justified in believing that *p*.

This definition retains clause (1) of Audi's definition, the justification clause, but it follows Shafer-Landau in specifying that "attentively considering" a self-evident proposition may be required, in addition to "adequately understanding" it, in order for one to be justified in believing it. My definition eliminates clause (2) of Audi's definition, the clause about knowledge. By removing this clause, I avoid controversies about the concept of knowledge, and I weaken the notion of self-evidence, removing a possible objection to the claim that the common sense principles are self-evident. Finally, my definition, like Audi's, but unlike Shafer-Landau's, stipulates that a self-evident proposition must be true. As Audi says, a self-evident proposition is, intuitively, one whose truth is somehow evident "in itself."²⁰

The important point is that adopting my definition rather than Audi's or Shafer-Landau's does not affect the plausibility of the claim that moral naturalism is incompatible with the self-evidence thesis. The claim that self-evident propositions are a priori justifiably believable is no more and no less plausible on my definition than it is on Audi's and Shafer-Landau's definitions.

Unfortunately, my definition does not eliminate some of the more serious difficulties with Audi's and Shafer-Landau's definitions. There are three important issues. To understand the definition, we need to understand what is to be meant by "adequately understanding and attentively considering" a proposition *p*, by this being "sufficient" for one to be justified in believing that *p*, and by this being sufficient for one to be "justified" in believing that *p*. I will consider these issues in reverse order.

First, Audi and Shafer-Landau intend adequately understanding (and attentively considering) a self-evident proposition to be sufficient merely for one to be justified *other things being equal* in believing it. Both of them

20 Audi 2004, 48. Shafer-Landau said in discussion (March 2006) that it was simply an oversight that he did not stipulate that self-evident propositions must be true.

hold that the justification possessed by self-evident beliefs is defeasible.²¹ I shall say, then, that adequately understanding and attentively considering a self-evident proposition is sufficient for one to be “defeasibly” justified in believing it.

Second, neither Audi nor Shafer-Landau argue that adequately understanding and attentively considering a self-evident proposition is *logically* or *strictly* sufficient for it to be the case that one is defeasibly justified in believing it. I think they intend to claim merely that if a proposition *p* is self-evident, then anyone in this world – or in a world with the same laws of nature – who had ordinary human cognitive capacities and who adequately understood *p* and attentively considered it would be defeasibly justified in believing that *p*.²² That is, I will say, adequately understanding and attentively considering a self-evident proposition is “ordinary-world sufficient” for defeasibly justified belief.

Finally, and most important, neither Audi nor Shafer-Landau say a great deal to clarify the idea of adequate understanding. Audi distinguishes “adequately understanding” a proposition from “mistaken or partial or clouded” understanding. He adds that “Adequacy here implies not only seeing what the proposition says but also being able to apply it to some appropriate cases, being able to see some of its logical implications, and comprehending its elements and some of their relations.”²³ He notes that there are cases in which adequately understanding a proposition requires “reflection” to clarify what it says.²⁴ He says very little more. Shafer-Landau says, in one place, that different conceptions of self-evidence will

21 Shafer-Landau 2003, 257–258; Audi 1997, 38.

22 Shafer-Landau points out that Audi defends the idea of relativizing self-evidence to agents’ capacities for understanding because propositions might be self-evident for a god that would not be self-evident for an ordinary human being. See Shafer-Landau 2003, 257 n. 7, citing Audi 1999, 222. This at least seems to open the door to the idea that a proposition might be self-evident to human beings in the actual world that would not be self-evident to beings in another world.

23 Audi 1997, 45; Audi 2004, 49–51.

24 Audi is not clear as to whether such reflection may involve inference. He says that if inference is involved, it is restricted “largely to clarifying what the proposition in question says.” See Audi 1997, 45–46. But on p. 102, he contrasts reflection with inference. A central claim of Audi’s intuitionism is that some moral beliefs do not need to have a basis in inference to count as justified or to count as knowledge. See Audi 2004, 25, 41. My focus here is on the idea of self-evidence, however, and it is not clear whether Audi thinks that a proposition that we can justifiably believe only after due reflection might still qualify as self-evident if the reflection involves inference. For discussion of Audi’s intuitionism, see Sinnott-Armstrong 2007.

explain it differently and that he wishes to remain neutral. In another place he says he views it as an open question whether one can adequately understand a moral proposition without having an appropriate accompanying motivation. He recommends Robert Audi's discussion as a "good starting point."²⁵

A key question is whether 'adequately understanding' a proposition *p* involves more than would be involved merely in meeting the minimal conditions required for 'grasping' *p*. The idea of 'grasping' a proposition is itself problematic, but I take it that to grasp a proposition *p* is to have the kind of cognitive acquaintance with *p* that is involved in believing that *p*. We cannot believe that *p* unless we have some minimal understanding of *p*, and it is this minimal understanding that I have in mind when I write about 'grasping' a proposition. It seems to me that both Audi and Shafer-Landau would insist that adequately understanding a proposition goes beyond merely grasping it. Audi says that adequately understanding a proposition implies more than merely seeing what the proposition says.²⁶ And he says there is a level of understanding a self-evident proposition that is, he seems to say, sufficient for believing it, but that falls short of adequate understanding, and that therefore is not sufficient for justification.²⁷ Shafer-Landau does not rule out construals of adequately understanding a proposition *p* that seem to go beyond requiring merely what would be involved in believing that *p*.²⁸ I take it, therefore, that more is required, in order to adequately understand a proposition, than merely to grasp it. Yet it is not clear what more is required. This is a crucial fuzziness in the definition.

We can see, nevertheless, that the definition is compatible with stronger and weaker accounts of self-evidence, depending on how we understand adequate understanding. At one extreme is an account according to which anyone who grasps a proposition adequately understands it; at the other extreme is an account according to which adequately understanding a proposition requires, as we might say, seeing all its ramifications. Different characterizations would set the threshold between self-evident and non-self-evident propositions at different places. In what follows, I will work with a vague and intuitive understanding of what is required, beyond grasping a proposition, in order to adequately understand it.

25 Shafer-Landau 2003, 249 n. 3 and 279 n. 12.

26 Audi 1997, 45.

27 Ibid., 110 n. 22.

28 Shafer-Landau 2003, 279 n. 12.

The issue, then, is whether moral naturalism is incompatible with the self-evidence thesis. Given the epistemic account of moral naturalism that I have proposed, the issue boils down to whether there are self-evident propositions about the instantiation of moral properties that are both synthetic and empirically indefeasible. A naturalist could deny that there are any self-evident moral propositions. Or she could deny that any self-evident moral propositions are synthetic. It seems plausible to me, however, that the common sense principles are both self-evident and synthetic. I therefore want to argue that they are empirically defeasible. They are synthetic and self-evident, but they are not strongly *a priori*.

As an example of a common sense principle, I will work with the proposition that, other things equal, it is wrong to take pleasure in another's pain. Call this "the pain proposition." Shafer-Landau claims that adequately understanding and attentively considering this proposition is sufficient for one to be defeasibly justified in believing it. This claim seems plausible to me, on an intuitive understanding of the relevant notions. So I will simply assume for the sake of argument that the pain proposition is self-evident. Moreover, although I cannot prove this, it seems plausible that the pain proposition and the other common sense principles are synthetic.²⁹ So I will simply concede for the sake of argument that they are.

One might object that the definition of a self-evident proposition ensures that all self-evident propositions are conceptual truths. But it does not, chiefly because, as I have assumed, 'adequately understanding' a proposition (in the sense in which this phrase is used in the definition) requires more than merely grasping it. On the familiar view, a proposition is a 'strict' conceptual truth only if a person who met the minimal conditions required for *grasping* it would thereby be in a position to know it merely by entertaining it and considering it carefully. But even if *p* is self-evident, so that adequately understanding and attentively considering *p* would be sufficient for one to be justified in believing that *p*, it might

29 On any realist view, the pain proposition is true only if there is a property of wrongness. Hence the pain proposition is a conceptual truth only if it is a conceptual truth that there is a property of wrongness. Yet I think it is not a conceptual truth that there is a property of wrongness. If it were, there would be a simple argument from a conceptual truth to the falsity of Mackie's "error theory." But even though I think Mackie's theory is false, I find it highly implausible that it is so easily shown to be false. For the error theory, see Mackie 1977, ch. 1. For a more detailed statement of this argument, see chapter 4.

not be the case that merely to grasp *p* and consider it carefully would be sufficient to enable one to know that *p*. It follows that *p* might not be a ‘strict’ conceptual truth. And it obviously does not follow from the definition alone that a self-evident proposition must be derivable from a set of conceptual truths if it is not itself a strict conceptual truth. It appears, then, that the definition of a self-evident proposition does not rule out the existence of synthetic self-evident moral propositions.

The idea that there can be synthetic self-evident propositions is puzzling. If the pain proposition is not a strict conceptual truth, it seems that one would need to take into account more than its content in order to be justified in believing it. It is therefore unclear how simply to understand it adequately, and attentively to consider it, could be sufficient to justify believing it.

5. SELF-EVIDENCE AND THE A PRIORI

I am assuming that the common sense principles are self-evident and that they are synthetic. We need to ask whether it follows from their being self-evident that they are *a priori*. Audi claims that the answer to this question is affirmative.³⁰

This claim is mistaken, however, because, again, as I am assuming, more may be required in order for a person to count as ‘adequately understanding’ a proposition than that she merely ‘grasp’ it. The traditional idea is that *a priori* knowledge and *a priori* warranted belief are ‘independent’ of experience – except for the experience required to understand the proposition in question (or that proposition and other propositions from which it follows). And I believe that the traditional idea is that we ‘understand’ a proposition in the relevant sense if we *grasp* it. But a person may need more experience in order to ‘adequately understand’ a proposition than merely to ‘grasp’ it. It therefore remains possible that a person who believes a self-evident proposition on the basis of adequately understanding and attentively considering it does not qualify as justified *a priori* in believing it.³¹

30 Audi appears to hold that all propositions that are self-evident in his sense are *a priori*. Audi 2004, 211 n. 19. See also p. 2, and see Audi 1997, 62 n. 24.

31 We could stipulate that a belief that *p* counts as ‘*quasi a priori* justified’ if it is justified exactly on the basis of ‘adequately understanding and attentively considering’ *p* – using these terms in the sense they are used in the definition of self-evidence. On this stipulation, any self-evident proposition is *quasi a priori*. Hence the common sense principles are *quasi a priori* on the assumption that they are self-evident. This is uninteresting.

But we need to be careful in formulating the key question. The central issue is whether the self-evidence thesis is compatible with the thesis that the moral properties are natural properties. Given my epistemic conception of a natural property, the key question is whether, if the common sense principles are self-evident, it follows that they are *strongly* a priori.

I do not think it even follows that they are *weakly* a priori. Let us assume that a person who is justified in believing a proposition is reasonable to believe it. On this assumption, Field's definition, as I believe he intended it to be read, implies that a proposition *p* is weakly a priori only if *grasping* *p* and considering *p* carefully are sufficient for a person to be justified in believing that *p* *even if* she has no empirical evidence that *p* is true other than whatever evidence she would need to have merely in order to *grasp* *p*. As I explained, I take it that adequately understanding a proposition in the sense of the term used in the definition of self-evidence requires more than merely grasping the proposition. Hence, it may be that in order to adequately understand *p*, a person would need to have evidence bearing on its truth that is not needed merely in order to grasp *p*. Therefore, even if a proposition *p* is self-evident, it does not follow that merely to grasp *p* and carefully consider it would be sufficient to justify believing that *p*. Therefore, even if the common sense principles are self-evident, they might not be weakly a priori.

The important point is that even if the common sense principles are self-evident, it does not follow that they are *strongly* a priori. For even if a proposition *p* is such that adequately understanding and attentively considering *p* is sufficient for one to be justified in believing that *p*, it does not follow that there can be no direct empirical evidence that not-*p*. Both Audi and Shafer-Landau acknowledge that the justification possessed by self-evident propositions is defeasible.³² Even if a person's belief that *p* could be defeasibly justified without empirical evidence that *p*, it does not follow that there can be no direct empirical evidence that not-*p*.

If I am correct, then even if the common sense principles are self-evident, if they are synthetic, they might well be empirically defeasible. It appears, then, that their being self-evident is compatible with moral naturalism. We are left, however, with two key questions. First, how can there be synthetic truths that are self-evident? Second, how could a naturalist make sense of the idea that the common sense principles are

32 Shafer-Landau 2003, 257–258; Audi 1997, 38; Audi 2004, 55.

self-evident even though they are synthetic and empirically defeasible? In the [next section](#), I aim to answer these questions.

6. A NATURALISTIC ACCOUNT OF SELF-EVIDENCE

I want to pursue the idea that in learning to understand certain propositions, we can come to be in a position to believe them justifiably, at least defeasibly. The key point is that a synthetic and empirically defeasible proposition might be such that in acquiring the concepts required to understand it we come to be in a position to be defeasibly justified in believing it. This explains how it could be that a synthetic and empirically defeasible proposition is self-evident.

Consider an example that Saul Kripke used in arguing that there are contingent a priori propositions. Kripke asks us to suppose that the term “meter” is introduced by reference to the length of S, which is a certain metal bar in Paris, and he then asks us to consider the proposition that S is one meter long. Kripke remarks that this proposition is contingent, since although S actually is one meter long, it could have been a different length. Moreover, Kripke suggests, the proposition is synthetic.³³ Yet, he claims, someone who has used S to fix the reference of “meter” knows a priori that S is one meter long, for she knows this “automatically, without further investigation.” In this sense, Kripke says, there are contingent a priori truths.³⁴ In the terms introduced by Field, if a person has used S to fix the reference of “meter,” the proposition that S is one meter long is weakly a priori. But it is not strongly a priori. Evidence that S is not in fact the bar that was used to fix the reference of “meter” and that it is in fact shorter than the standard meter bar would be evidence that the proposition is false.³⁵ So, plausibly, the proposition is weakly a priori, yet synthetic and empirically defeasible.

I shall propose an argument from concept acquisition. The idea is that, given certain assumptions about how we acquire a concept, certain propositions ‘involving’ the concept can be weakly a priori yet synthetic and empirically defeasible. By exploiting this idea, I believe a naturalist can explain how the common sense moral principles could be self-evident even though they are synthetic and empirically defeasible.

33 Kripke 1980, 56 n. 21. To be exact, Kripke says he “chooses” not to call the proposition “analytic” (39 and 122 n. 63).

34 Ibid., 54–57, esp. 56.

35 It is reasonably clear that Kripke would agree about this. Ibid., 54–57.

Suppose that we acquire the concept of gold by being shown some things, such as a wedding ring and a gold pocket watch, and by being told that gold is the malleable yellow metal with which these things were made. Kripke remarks in a similar context that it could turn out that some of the things are not actually gold; some of them might be fool's gold.³⁶ Nevertheless, it seems to me, a person who acquired the concept of gold in something like the way I imagined would be defeasibly justified, "without further investigation," in thinking that gold is a yellow metal. Clearly, however, the proposition that gold is yellow is contingent and empirically defeasible. Kripke points out that there could be evidence that the yellow appearance of gold is due to an optical illusion.³⁷ So, it seems, the proposition that gold is yellow is weakly a priori yet synthetic and empirically defeasible.

Assume, then, that we acquire the concept of wrongness by being given paradigmatic examples of kinds of action and by being told that these kinds of action are wrong, at least other things being equal. Suppose in particular that, in acquiring the concept, we are taught that it would be wrong to take pleasure in another's pain, to taunt and threaten the vulnerable, to sell another's secrets solely for personal gain, to torture others just for fun, and so on. A person who acquired the concept in roughly the way I have assumed would be justified, "without further investigation," yet defeasibly, in believing, for example, that it is wrong other things being equal to take pleasure in another's pain. For such a person, the pain proposition would be weakly a priori. Moreover, if the proposition is true, it qualifies as self-evident. For if a person learns to understand the pain proposition through, among other things, acquiring the concept of wrongness in something like the way I have described, then she would be defeasibly justified in believing it. If she was taught what wrong action is by being taught, *inter alia*, that it is wrong other things being equal to take pleasure in another's pain, then if she understood what she was taught, she would be defeasibly justified in believing that it is wrong other things being equal to take pleasure in another's pain. Hence, given that the pain proposition is true, and given our assumption about how we acquire the concept of wrongness, the proposition qualifies as self-evident.

I conclude, then, that the intuition that the common sense principles are self-evident can be explained by our hypothesis about how we acquire

36 Ibid., 135–136.

37 Ibid., 118.

the concept of wrongness. It is plausible that a person who adequately understood any one of the common sense principles would be in a position to believe it justifiably, and it is also plausible, and for the same reason, that if a person denied it, we might wonder whether she understood what wrongness is. This can be explained by our assumption about how we acquire the concept of wrongness. For if a person learns the concept of wrongness by learning, among other things, that the kinds of action referred to in the common sense principles are paradigmatic examples of wrong actions, then she would be defeasibly justified to believe the principles. The principles would qualify as self-evident provided they are true. The argument for this would be the same, *mutatis mutandis*, as the argument I gave above regarding the pain proposition.

No doubt my assumption about how we acquire the concept of wrongness is unrealistic and oversimplified. Moral teaching need not be as explicit as I am assuming it to be. However, my objective is merely to show that moral naturalism is compatible with the thesis that there are synthetic self-evident moral principles, so I think it is fair to begin with an idealized picture. Yet there are complexities that must be acknowledged.

Most important, even if the concept of wrongness is taught by means of paradigmatic examples of kinds of action that are wrong, it is likely that different examples would be used in different contexts, and people who disagree about what kinds of actions are wrong might reject each other's examples. Given the account of moral concept acquisition I am working with, this might seem to imply that there is not a single concept of wrongness. But I think not. It is plausible that people acquire the concept of gold in part by being given certain examples, yet there is not a fixed set of samples that is used in every context by everyone teaching the concept, and there can be disagreement about some samples. We nevertheless manage to acquire the same concept, with due allowance for vagueness.³⁸ Similarly, in the moral case, it seems to me, even if the concept of wrongness is taught by means of different examples of kinds of action that are wrong, we manage to acquire the same concept, with due allowance for vagueness.³⁹

38 Kripke comments on vagueness in the idea of 'gold.' Ibid., 136.

39 Obviously there are circumstances in which we would think that people had not acquired the same concept of wrongness even if they use what seems to be the same word. Once we see this, we might begin to have worries reminiscent of the worries that underlie the 'Moral Twin Earth' argument. I discuss this argument in chapters 6 and 7.

It turns out, however, on the account I have given, that propositions are not self-evident simpliciter, but rather, self-evidence must be relativized to believers' cognitive situations.⁴⁰ Suppose, for example, that a person acquired the concept of wrongness in part by being given K actions as paradigmatic examples of wrong actions. In this case, the proposition that K actions are wrong might be self-evident-in-relation-to the cognitive situation of this person. This person would be defeasibly justified in believing that K actions are wrong simply in virtue of attentively considering the proposition that K actions are wrong, given how she acquired the concept of wrongness. We therefore need to revise our definition. We should say this:

A proposition *p* is self-evident in relation to a cognitive situation *S* just in case *p* is a truth such that a person in *S* who adequately understood and attentively considered just *p* would be defeasibly justified in believing that *p*.

One might object that too many propositions qualify as self-evident on this proposal.⁴¹ Any true moral principle that ascribes a moral property to something *x* would qualify as self-evident relative to the cognitive situation of someone who acquired the concept of that property by being given *x* as a paradigmatic instance of something with the property. Yet I do not see this as a problem. A moral principle will qualify as self-evident in relation to an epistemic situation only if it is true and only if it seems platitudinous to people in that situation.

The argument from concept acquisition is one way that a naturalist might attempt to explain the thesis that the common sense principles are synthetic yet self-evident. Perhaps there are other ways. The important point is that the argument is compatible with the thesis that the common sense principles are empirically defeasible. Of course, it will not be clear what kind of evidence could tell against their truth unless we examine a specific naturalistic theory. Different theories will have different things to say.

40 Consider the negation of the pain proposition, the 'sadists' principle' that it is permissible to take pleasure in another's pain. A community of sadists in which the sadists' principle is treated as a platitude is not a cognitive situation in which the pain proposition is self-evident. But, of course, since the sadists' principle is false, it is not self-evident.

41 The proposition that fire engines are red might qualify as self-evident in relation to the cognitive situation of a person who acquired the concept of redness in part by being given the color of fire engines as a paradigmatic example of redness. I owe this objection to Jon Tresan, in conversation.

7. THE EMPIRICAL DEFEASIBILITY OF SELF-EVIDENT MORAL PRINCIPLES

To explain how a self-evident moral principle could be empirically defeasible, it will help to work with a specific version of moral naturalism. I will sketch such a theory, the *society-centered* moral theory that I have proposed and defended in detail elsewhere.⁴²

The central idea, roughly speaking, is that a basic moral *proposition* is true only if a corresponding moral *standard* has a relevantly authoritative status.⁴³ By a “standard,” I mean a content expressible by an imperative. For example, it is wrong to take pleasure in another’s pain only if (roughly) a standard prohibiting people from allowing themselves to take pleasure in another’s pain has the relevant status.⁴⁴ A moral standard has the relevant status just in case (roughly) its currency in the relevant society would best contribute to the society’s ability to meet its needs – including its needs for physical continuity, internal harmony and cooperation, and peaceful and cooperative relations with its neighbors. This view treats moral properties as relational. If torture is wrong, it is wrong in relation to a given society, a society in which the currency of a standard prohibiting torture would best contribute to the society’s ability to meet its needs. We can think of the moral standards with currency in a society as forming a social moral code. Moral claims are true or false depending on the content of the moral code the currency of which in the relevant society S would best contribute to S’s ability to meet its needs. Call this the “S-ideal code.”

Society-centered theory raises a number of difficult questions, including questions about its semantics and its metaphysics of moral properties. For example, which society is the relevant society in a given context?⁴⁵

42 Copp 1995. For a brief introduction to the view, see Copp 1996.

43 A basic moral proposition is such that, for some moral property M, it entails that something instantiates M. An example is the proposition that capital punishment is wrong. Among nonbasic moral propositions are propositions such as that nothing is morally wrong and that either abortion is wrong or $2 + 2 = 4$. In Copp 1995, I called basic moral propositions “paradigmatic.”

44 In a fuller discussion of society-centered theory, I would qualify this claim. It is correct that a basic moral claim is true only if a corresponding standard has the relevant status, but there are other conditions necessary for the truth of some moral claims. Some details are set out in Copp 1995, e.g., 24–26, 28–30.

45 On the account I proposed in Copp 1995 (221), this is settled by the pragmatics of moral conversation and thought. In the default case, the society at issue in a given context is the smallest society that embraces all the people referred to in the context, all the people the speakers or thinkers have in mind in the context, and all the people who are party to the conversation.

This is not the place to go into detail, nor to attempt to explain why I find the theory plausible. I merely want to let the theory serve as an example.

The theory implies that moral properties are natural properties in the sense I have explained. It implies that the predicate “wrong,” when used in a given context, ascribes, roughly, the property of being ruled out by the S-ideal code, whatever it actually is, where S is the society that is relevant in the context. If M is actually the S-ideal code, then an act is wrong (in relation to S) only if it is ruled out by a norm that is entailed by or included in M. Call wrongness-in-relation-to-S “S-wrongness.” I take it to be obvious that any substantive belief we had about this property would be empirically defeasible. Evidence that the S-ideal moral code would not in fact contain an injunction against torture would be evidence that torture is not S-wrong, and so, on the theory, it would be evidence that the proposition expressed by “Torture is wrong” in a context where S is the relevant society is not in fact true. Hence, the proposition expressed by “Torture is wrong” in any given context is not strongly a priori. According to the theory, no synthetic moral proposition would be strongly a priori. According to the theory, then, even if the common sense principles are self-evident, they are empirically defeasible.

8. CONCLUSION

My goal in this chapter has been to show that moral naturalism is compatible with the idea that there are synthetic self-evident moral truths such as the common sense principles. According to my proposed account of the idea of a natural property, moral naturalism is best understood to combine moral realism with the denial, roughly, that there are synthetic, strongly a priori, moral propositions. Since it is plausible that the common sense principles are synthetic, and since it is plausible that they are self-evident, the issue for the naturalist boils down to whether they are strongly a priori. I argued that it is compatible with moral naturalism to hold that the common sense principles are self-evident, at least in certain cognitive contexts, for holding this about the principles is compatible with claiming that the principles are empirically defeasible. A naturalist therefore can agree that there are synthetic moral principles that are self-evident.

REFERENCES

- Audi, Robert. 1996. “Intuitionism, Pluralism, and the Foundations of Ethics.” In Sinnott-Armstrong and Timmons 1996: 101–136.
- . 1997. *Moral Knowledge and Ethical Character*. New York: Oxford University Press.

- . 1999. "Self-Evidence." *Philosophical Perspectives* 13: 205–226.
- . 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In Sinnott-Armstrong and Timmons 1996: 243–266.
- Dancy, Jonathan. 2004. *Ethics Without Principles*. Oxford: Oxford University Press.
- Field, Hartry. 2000. "Apriority as an Evaluative Notion." In *New Essays on the A Priori*, ed. Paul Boghossian and Christopher Peacocke, 117–149. Oxford: Clarendon Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Clarendon Press.
- Sinnott-Armstrong, Walter. 2007. "Reflections on Reflection in Robert Audi's Moral Intuitionism." In *Rationality and the Good*, ed. Mark Timmons, John Greco, and Al Mele. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter, and Mark Timmons, eds. 1996. *Moral Knowledge? New Readings in Moral Epistemology*. New York: Oxford University Press.

Moral Necessities in a Contingent World

1. INTRODUCTION

Certain moral propositions seem intuitively to be necessarily true. We think it would be wrong, other things being equal, to cause someone pain for no reason, to take pleasure in another's pain, to punish those known to be innocent, to taunt the vulnerable.¹ We think it would be wrong to torture a baby to death just for fun.² These propositions seem to be necessarily true – in some interesting sense of “necessary.” They seem to be “moral necessities,” as I will say. We do not countenance possible worlds in which they are false – except, perhaps, worlds that are radically different from ours in some crucial respect. We can call the worlds in which we take these propositions to be true the “morally relevant worlds” or “M-worlds” – and we can say that we view the propositions as “M-necessarily” true.

The moral principles proposed by normative moral theories are typically offered and treated as putative necessary truths. They are supposed to be immune to counter-examples, and it is standardly assumed that a counter-example need only be a possible case in which the principle implies something that is intuitively false. Consider, for example, a

I am grateful to Michael Jubien and Thomas Hofweber for helpful discussion about the nature of properties and about modality, and to Richard Kraut, David McNaughton, Philip Pettit, Henry Richardson, John Skorupski, and, especially, Stephen Finlay, Michael Ridge, Walter Sinnott-Armstrong, and Jon Tresan, for extensive written comments and suggestions. I presented versions of this chapter to the 32nd Conference on Value Inquiry, at Louisiana State University, in April 2005, and to the 4th Conference on Moral Theory and Practice, in Granada, Spain, in June 2005. I am grateful to both audiences for valuable discussion.

1 All but the first of the examples come from Shafer-Landau 2003, 248.

2 This example is from Thomson 1990, 18.

classical act-consequentialist principle according to which an agent is morally required to perform an act with consequences that are at least as good as the consequences of any alternative. If this principle is correct, then, it is standardly assumed, it is not merely true in the actual world, but it is true in all relevant worlds – let us again call these the “M-worlds.” This is why it counts as an objection if one can describe a possible case in which the right thing to do seems to be something that would produce less good than some alternative. A consequentialist has a variety of familiar replies to such objections, of course. However, an admission that the principle is only true in the actual world, or that it is false in some intuitively relevant world, would standardly be taken as an admission that the principle is incorrect. This suggests that the standard view is that the principle is correct just in case it is M-necessary.

The thesis that there are moral necessities is nevertheless puzzling. It is plausible to think that what morality requires of us depends on human nature and on the circumstances of human life. But these are contingent matters, so it may seem difficult to understand how there could be any interesting moral propositions that are necessarily true. Moreover, the examples with which I began are substantive moral claims that appear not to be conceptual truths, and it is unclear how such a claim could be necessarily true. Hence, the thesis that there are moral necessities raises a metaphysical question: If there are moral necessities, what grounds their necessity? This is the question that I will be addressing.

I believe that an adequate metaethical theory that acknowledges the existence of moral necessities must explain the nature and grounding of the necessity involved. The idea that metaethics faces this constraint – “the moral necessities constraint” – is supported by the idea that an adequate metaphysics faces a similar constraint – the need to explain the nature and grounding of any necessities that it acknowledges.³ The constraint is a corollary (so to speak) of the job description of metaethics. For one major task of a metaethical theory is to provide an account of the truth conditions of moral propositions – assuming, that is, that the theory is cognitivist, so that it acknowledges that there are moral propositions – and an account of the truth conditions of moral propositions would yield an account of the truth conditions of putative moral necessities. To be sure, some philosophers would reject the constraint. They might have minimal explanatory goals and be willing to take moral necessities as unexplained

3 Michael Jubien presses the idea that there is such a requirement in Jubien 2007.

primitives. Yet even they can join me in the project of this chapter, for my project is to explore issues that arise if we attempt to explain the moral necessities.

My central thesis is that only certain kinds of naturalistic moral realism are capable of adequately explaining the nature and grounding of moral necessities. This may seem surprising, for, at least as I understand things, a naturalist holds that any substantive moral knowledge we have is empirical,⁴ and it may be difficult to believe that we could have empirical knowledge that a substantive moral claim is necessarily true. I shall be arguing, however, that certain kinds of naturalism are able to ground moral necessities. If we believe there are moral necessities, then unless we are prepared either to accept an antirealist metaethics, such as would be proposed by noncognitivism, or to reject the moral necessities constraint, the fact that these theories are able to explain the moral necessities is a reason to favor them.

After clarifying the thesis that there are moral necessities, I explore the possibility of explaining them as conceptual necessities and as metaphysical necessities. I argue that neither approach is viable – not unless a kind of reduction or analysis of moral properties can be defended. Since moral nonnaturalism is typically antireductionist, the options I explore are versions of moral naturalism. I argue that there are kinds of naturalism that can account for the moral necessities and explain their status as metaphysical necessities.⁵ At the end of the chapter, I explore a Kantian option and a noncognitivist option. I argue, even if all too briefly, that neither provides a satisfying account of moral necessities. The upshot is, I hold, that there is reason to favor moral naturalism.

2. THE THESIS THAT THERE ARE MORAL NECESSITIES

Let me stipulate for now that a moral proposition is a “moral necessity” just in case it is necessarily true – or M-necessarily true. Later in this section I will say more about the notion of M-necessity. If a proposition *p* is a moral necessity, then I will also speak of the proposition that it is

4 I explain this in chapter 1.

5 Kit Fine argues that moral necessities are neither conceptual nor metaphysical and that, in order to accommodate them, we must postulate the existence of *sui generis* normative necessities. He reaches this conclusion after considering the naturalistic option, but rejecting it for independent reasons. See Fine 2002. For discussion of related issues, see Weatherson 2004. I do not want to postulate an unexplained and primitive kind of normative necessity.

M-necessary that *p* as a moral necessity. The context should make clear which kind of proposition is at issue.

I will work with an example that seems to be especially plausible, a proposition I will call the “painful death proposition.” It is the proposition that it would be wrong (other things being equal) intentionally to cause a child to die a prolonged and excruciatingly painful death – or to cause a child a “painful death,” as I shall say for the sake of brevity. I will call the property an action can have of being wrong, *other things being equal*, the “wrongness property.” Accordingly, I might say that the painful death proposition ascribes the wrongness property to actions that cause a child a painful death.

The moral necessities that are at issue in this chapter contrast with contingent moral truths, such as the proposition that Alice is morally required to visit Brenda in the hospital. This proposition is contingent since there is a morally relevant possible world in which Brenda has no need of a visit because she is a doctor rather than a patient needing solace.⁶

Some moral necessities are trivial and uninteresting. Consider, for example, the proposition that murder is wrong. It is at least arguable that, given the meaning of the term “murder,” the sentence “Murder is wrong” is analytic. If this sentence is analytic, moreover, then it is necessary that murder is wrong, for an action would not qualify as a case of murder unless it were wrong. This is uninteresting and unsurprising. Perhaps it will be denied that “Murder is wrong” is analytic, but if so, we can stipulate that a “schmurder” is a wrongful killing of a human being. Given this stipulation, the sentence “Schmurder is wrong” is analytic, and it is a moral necessity that schmurder is wrong.⁷ This is the same uninteresting

6 There are two possible sources of confusion here. First, there is a familiar Kantian usage in which any moral requirement may be called a “moral necessity.” This usage is found, for example, in Darwall 2006. If Alice is morally required to visit Brenda in the hospital, there is a sense in which it is necessary that she do it; she must do it. Yet of course it is contingent that she must do it. We therefore need to distinguish the moral necessities that are at issue in this chapter from “moral necessities” in the Kantian sense. Second, deontic logic treats propositions of the form (It is morally obligatory that *p*) as modal propositions formally analogous to propositions of the form (It is necessary that *p*). This way of handling the semantics of moral language has certain formal advantages, since it permits drawing on the resources of modal logic. But “It is morally obligatory that Alice visit Brenda in the hospital” does not express a proposition that is a moral necessity in the sense at issue here.

7 Quinean worries about analyticity cannot rule out the possibility of stipulating that “schmurder” means “wrongful killing of a human being.” On this stipulation, the proposition expressed by “Schmurder is wrong” is the proposition that the wrongful killing of a human being is wrong. Quinean doubts about analyticity in natural languages give us no reason to deny that this proposition is a moral necessity. See Quine 1961.

moral necessity that, if I am correct, we could express in ordinary English by saying that murder is wrong.⁸

Unlike the uninteresting proposition that schmurder is wrong, the examples with which I began this chapter appear to be substantive. We can explain a sense in which they are substantive by exploiting J. L. Mackie's "error theory." According to this theory, there are no moral properties, and hence, no actions are wrong and no actions could be wrong.⁹ It follows that there could not be any wrongful killings, and so, given the above stipulation, it follows in turn that there could not be any cases of schmurder. Nevertheless, given the above stipulation, Mackie can agree that any case of schmurder would be wrong. The important point is that Mackie is forced by the error theory to deny the painful death proposition. He must admit, of course, that there could be a case in which a person causes a child to die a prolonged and excruciatingly painful death. But the error theory entails that the person's action in such a case would not be wrong, not even if we assume that 'other things are equal.'¹⁰ Hence,

8 One might think that act-consequentialists would deny that it is necessarily the case that schmurder is wrong. For, on their view, the wrongness of schmurder, like anything else, must depend on its consequences. However, given the stipulated meaning of "schmurder," act-consequentialism entails that a killing is not a schmurder unless it is a case of killing a human being where the agent had available an alternative that would have had better consequences. Hence, the act-consequentialist need not deny that it is a moral necessity that schmurder is wrong.

9 See Mackie 1977, ch. 1. I take the error theory to consist of two claims: First, there are no moral properties, and second, because of this, any proposition that putatively ascribes a moral property to something is false. Moreover, I assume that if a property exists, it exists necessarily. Hence, it follows from the theory that nothing could have a moral property. So, for example, any claim to the effect that something is or could be wrong is false. There are problems in the interpretation of the theory. On certain views, if there is no property ascribed by "wrong," then sentences such as "Lying is wrong" would not even express propositions. On these theories, the error theory would commit Mackie to viewing such sentences as meaningless. However, Mackie does not take this view. He takes the error theory to imply that claims such as that lying is wrong are false, not that sentences such as "Lying is wrong" fail to express propositions and are meaningless. I assume that a plausible semantics would provide a way of understanding Mackie's theory that would avoid this problem. An alternative reading of the theory might take it to say that there is a property of wrongness but that it could not possibly be instantiated. This does not seem to have been Mackie's view, however.

10 One might object that it is compatible with the error theory to hold that the painful death proposition is true provided we hold it is not the case that 'other things are equal.' For if other things are not equal, then the painful death proposition does not ascribe wrongness to anything. But I take the idea that it would be wrong 'other things being equal' to cause a child to undergo a painful death to be the idea that doing such a thing *would* be wrong *period* in any situation in which there was no moral reason or justification for causing a child to undergo a painful death. And the error theory entails that there are no moral reasons

Mackie must deny that it would be wrong other things being equal to cause a child to die a prolonged and excruciatingly painful death. The upshot is that he does *not* deny the tautological moral necessity that any schmurder would be wrong. However, he *does* deny the substantive claim that it would be wrong other things being equal to cause a child to die a prolonged and excruciatingly painful death. This suggests an informal “Mackie test” of whether a moral necessity is substantive. If Mackie is committed to denying a moral necessity by the error theory (together with what he must admit, on independent grounds, might happen), it is substantive – “Mackie-substantive” – otherwise it is not.

In this chapter, I shall be discussing the thesis that there are *substantive* moral necessities: Mackie-substantive moral necessities. It is worth mentioning, however, that moral necessities that are not Mackie-substantive can be philosophically important. It is widely agreed to be a necessary truth that moral facts supervene on nonmoral facts. The basic idea is that if two situations differ morally, then they differ in some nonmoral way (in addition to any differences there may be between them as to which particulars are involved in them). This conditional is a moral necessity, or so I assume, yet Mackie’s error theory does not commit him to denying it, so it is not Mackie-substantive. Despite this, however, it is arguable that important metaethical consequences follow from it, or at least from some more precise formulation of it.¹¹

Let me return to a consideration of the painful death proposition, the proposition that it would be wrong, other things being equal, intentionally to cause a child to die a prolonged and excruciatingly painful death. It is plausible, I think, that this proposition is necessarily true – in the relevant sense of “necessary.” The idea that it would be wrong ‘other things being equal’ to cause a child to undergo a painful death is – I assume for present purposes – the idea that this *would* be wrong *period*, or all things considered, in the absence of a moral reason or justification for causing such a death.¹²

or moral justifications. Hence, on the error theory, other things are ‘always’ equal in the relevant sense, and so the proposition that it would be wrong other things being equal to cause a child to undergo a painful death has the same truth conditions as the proposition that it would be wrong period to cause a child to undergo a painful death. And on the error theory, the latter is false.

11 See Jackson 1998, 118–119. As Jackson formulates the supervenience thesis, it is the thesis that, roughly, if possible world *w* and possible world *w*’ differ morally, then they differ in some nonmoral way (in addition to any differences there may be between them as to which particulars are involved in them). He draws important conclusions from the thesis (*ibid.*, 119, 122–123).

12 See Kagan 1998, 180–181.

We can imagine extreme circumstances in which there would be a such a justification. Perhaps, for example, unless we cause a child to undergo a prolonged and excruciatingly painful death, every other human being will immediately undergo such a death. Yet even in circumstances of this kind, there remains a very strong reason *not* to cause a child a painful death, a reason that *would* make it wrong *period* to do this if the circumstances were more ordinary. Given this, the painful death proposition can perhaps be spelled out as follows: There is a moral reason not to cause a child a painful death and, in the absence of a reason or justification for causing a child a painful death, this reason would be sufficient to make it wrong *period* to cause a child to undergo such a death. To deny that this proposition is necessarily true, a person would have to think there are possible and relevant circumstances in which the fact that an action would cause a child a painful death is no moral reason at all against it or, at least, is not a sufficiently strong reason against causing a child a painful death that it needs to be counterbalanced by a reason or justification for doing so. To think this, it seems, one would have to be morally beyond the pale or, like Mackie, to be in the grip of a kind of moral nihilism.¹³

It might seem that a particularist would deny that the painful death proposition is a moral necessity. A particularist thinks that a consideration can be a moral reason *against* an action in one circumstance, even if a consideration of the same kind is a reason *in favor* of an action in another circumstance. Yet even a particularist could agree that there is always a moral reason against causing a child to undergo a prolonged and excruciatingly painful death.¹⁴ It might seem that a classical act-consequentialist would deny that the painful death proposition is a necessary truth, for there are possible circumstances in which someone's causing a child a painful death would have better consequences than anything else she

13 Or one might not understand the kind of death at issue. For a graphic description of such a death, see the account of the torture and execution of Damiens, "the regicide," in Foucault 1979, 3–5. I owe this reference to David McNaughton. Readers who doubt the necessity of the painful death proposition might want to consider instead the proposition that it would be wrong other things being equal to intentionally cause a child to undergo a prolonged and painful death of the kind experienced by Damiens.

14 For a defense of particularism, see Dancy 2004. Dancy allows for the possibility of moral considerations that are "invariant" in that they either always provide reasons against action or always provide reasons for action (78); he considers the example of causing gratuitous pain to unwilling victims (77). He insists, however, that such invariance, "where it occurs, derives not from the fact that we are dealing here with a reason, but from the particular content of that reason" (77). I am grateful to David McNaughton for helpful discussion of this point (in correspondence).

could do. Yet, plausibly, ‘other things are not equal’ in such circumstances, so the act-consequentialist can agree that it necessarily would be wrong *other things being equal* to cause a child to undergo a painful death. Of course, a moral nihilist would reject the necessity of the painful death proposition. A discussion of nihilism is, however, beyond the scope of this chapter.

One might worry that there are possible worlds in which nothing is right or wrong. Hume thought there could be situations in which ‘the circumstances of justice’ do not obtain, such that, in those worlds, nothing that a person does is just or unjust.¹⁵ We might then countenance possible worlds in which ‘the circumstances of morality’ do not obtain, such that nothing is right or wrong, and no moral property is instantiated. Perhaps moral requirements have certain presuppositions that simply do not hold in these worlds. We can set this worry aside, however, for the painful death proposition is plausibly taken to be true in all worlds that *are* in the circumstances of morality, including, I shall assume, the actual world. These are the worlds that I had in mind before when I introduced the idea of ‘morally relevant worlds’ or ‘M-worlds.’¹⁶ The painful death proposition is true in all such worlds. It is M-necessary.

From this point on, then, let us say that a moral proposition is a ‘moral necessity’ just in case it is true in all morally relevant possible worlds. Let it be understood, then, that a moral necessity would be ‘M-necessary.’ The thesis that interests us is that there are substantive moral necessities, so understood. In what follows, I will mostly ignore the distinction between M-necessity and necessity in the ‘strongest sense,’¹⁷ the necessity of a proposition that is true in all possible worlds.

Although I have been writing and shall continue to write of possible worlds, I take it that there are not *literally* any nonactual worlds. Talk of possible worlds is metaphorical. Hence, we do not *explain* the necessity of a proposition by claiming it to be true in all possible worlds, we merely say in a metaphorical way that it is necessary. Michael Jubien argues that necessities can be explained on the basis of certain kinds of relations between properties, such as the relation between a property and its constituents.¹⁸ It is plausible to think, for example, that the fact that it is necessarily the

15 Hume 1888, III, ii, 2; Rawls 1971, sec. 22.

16 We assume that the actual world is an M-world, for we want it to be the case that if a proposition is M-necessary, it is true in the actual world.

17 Fine writes of necessity in the “strongest sense” in Fine 2002.

18 See Jubien 2007, where he makes a detailed case against the literal existence of possible worlds.

case that a bachelor is unmarried is due to the fact that the property of being a bachelor is partly constituted by the property of being unmarried. I believe that this is a promising strategy and I shall exploit it freely in my argument.

It is typical to distinguish two kinds of propositions that are necessary in the strongest sense.¹⁹ First are ‘conceptual truths.’ An example is the proposition that a vixen is a fox, which we take to be necessarily true in virtue of the relation between the ‘concept’ of a vixen and the ‘concept’ of a fox. Second are ‘metaphysical necessities,’ the necessity of which is (putatively) grounded in some way in the nature of things.²⁰ An example is, perhaps, the proposition that any quantity of water is a quantity of H₂O. The issue for us is whether moral necessities can be explained in either of these ways. I begin with the idea of a conceptual necessity.

3. SUBSTANTIVE MORAL NECESSITIES AS CONCEPTUAL TRUTHS

On the familiar standard view, a proposition is a conceptual truth just in case the concepts that constitute it are related to one another in such a way that its truth is logically guaranteed. Moreover, because of this, a person need only grasp the proposition and consider it carefully in order to be in a position to know that it is true. I do not want to commit myself to the idea that propositions are constituted by concepts, however. I shall therefore attempt to explicate the idea of a conceptual truth without referring to concepts.

Begin with the idea that a proposition is a conceptual truth just in case any person who met the minimal conditions required for grasping it would thereby be in a position to know that it is true merely by considering carefully whether it is true while thinking clearly – merely by ‘entertaining’ it, as I shall say. That is, such a person would be in a position to know that it is true without acquiring any (further) empirical information and without drawing on any additional knowledge that she has. Of course, if the proposition is especially complex, she might grasp it but be unable to work out that it is true. The idea is, however, that if she entertained it, she would be ‘in a position’ to know that it is true.

This account presents us with a puzzle. How is it that, merely by grasping and entertaining a proposition, a person could be in a position

19 See Fine 2002.

20 Fine says metaphysical necessities are grounded in the “identity” of things. *Ibid.*, 254.

to know that it is true? It seems that the answer must be that the truth conditions of a conceptual truth are ‘internal to its content’ in the sense that its truth does not depend on anything but intrinsic properties of the proposition, properties that anyone who grasps the proposition is in a position to grasp.²¹ I will call these properties “intrinsic truth-ensuring properties,” and I will speak of such a proposition as having “internal truth conditions.”²² If conceptual truths have such properties, this would explain how a person who grasps and entertains a conceptual truth is thereby in a position to see that it is true. For the fact that the proposition has these properties would entail that it is true, where all of this would be accessible to anyone who grasps the proposition.²³ The proposal is, then, that conceptual truths have a metaphysical characteristic that underwrites their a priori epistemology. In particular, a proposition *p* is a conceptual truth only if *p* has certain intrinsic properties that one must be in a position to grasp, and to grasp that *p* has, in virtue of meeting the minimal

- 21 Saul Kripke argues that there are contingent a priori truths, and one might think that if Kripke is correct, the propositions he has in mind would be conceptual truths. Hence, we might need to allow that there are contingent conceptual truths. Fortunately, none of the examples I consider is plausibly taken to be a contingent conceptual truth, so I shall set the issue aside. See Kripke 1980, 54–57. One of Kripke’s examples is the proposition that *S* is one meter long, where *S* is the standard meter bar in Paris that has been used by someone to fix the reference of “meter.” He remarks that this person knows a priori that *S* is one meter long (56). The person knows a priori that *S* is one meter long because he used *S* to fix the reference of the term “one meter,” and as a result, the concept of a meter is fixed for him by reference to the length of *S*. There are, nevertheless, other ways to fix the concept of a meter, so it is not the case that anyone who grasped the proposition that *S* is one meter long would be in a position to know it is true merely by entertaining it. The proposition is therefore not a conceptual truth on the account I gave in the text. It is worth mentioning in this context that Jonathan Dancy has proposed that all substantive moral knowledge is contingent a priori (2004, 146–148). I owe this reference to Michael Ridge. I am grateful to Michael Jubien, Greg Ray, Michael Ridge, and Jon Tresan for helpful discussion of the contingent a priori.
- 22 I intend this account to be neutral among different accounts of the nature of propositions, provided that the accounts allow us to make sense of conceptual truths. I stipulate that a property is an ‘intrinsic truth-ensuring property’ of a proposition *p* just in case it is a member of the set containing properties of *p* (excluding the property of being true, if there is such a property) that are intrinsic to *p* and that are such that the fact that *p* has the properties in the set entails that it is true (and where each property in the set is essential to there being this entailment).
- 23 The proposition that there is this entailment must also be a conceptual truth. That is, if a proposition *p* is a conceptual truth, then it has certain intrinsic properties such that it is a conceptual truth that the fact that it has these properties entails that its truth conditions are satisfied. The fact that I use the notion of a conceptual truth in my account means, unfortunately, that my explanation of the a priori epistemology of conceptual truths depends on the very notion of a conceptual truth. I do not see how to avoid this.

conditions necessary for grasping p , where p 's having these properties entails that it is true, and where one must also be in a position to grasp this fact in virtue of meeting the minimal conditions for grasping p .²⁴ This is an account of 'strict' conceptual truths. I will soon explain the idea of a 'derivative' conceptual truth.

An example of a strict conceptual truth is the proposition that a vixen is a fox. This proposition is true only if there is a certain relation between vixens and foxes, a relation that must obtain given that the property of being a vixen includes or entails the property of being a fox. A person has not grasped the proposition unless she understands this. The proposition that a vixen is in the garden obviously is not a conceptual truth, however. For a person who merely understands enough about vixens and gardens to grasp the proposition need not be in a position to know where vixens are to be found.

Consider now propositions that do not have internal truth conditions, so that they are not strict conceptual truths, but each of which is entailed by a set of strict conceptual truths, where the proposition that there is this entailment is also a strict conceptual truth. I will call propositions of this kind "derivative conceptual truths." They would be knowable simply on the basis of grasping and entertaining certain other propositions with internal truth conditions and working out their entailments. They would have a status analogous to the status of theorems in an axiomatic system. When I want to refer to propositions of this kind as well as to strict conceptual truths, I will speak of conceptual truths "in the wide sense."

The proposal we are considering, then, is that the substantive moral necessities are conceptual truths in the wide sense.²⁵ A proposition p

24 One could say that on my account a concept is an intrinsic property of a proposition, a property that one must grasp in order to grasp the proposition. Hence the concept of a tree is a property that is shared by all propositions about trees, and that is such that one must grasp it in order to grasp these propositions. On this showing, the idea of grasping a proposition is prior in the order of explanation to the idea of a concept.

25 The paradox of analysis challenges the idea that there can be substantive conceptual truths. For discussions of the paradox, see Smith 1994 and King 1998. A familiar response to the paradox is that a conceptual truth may be substantive in the sense that it may be difficult to grasp or to entertain, or it may be difficult for a person who grasps and entertains it to work out that it is true. This is a sense of "substantive" different from that isolated by the Mackie test: There could be a moral necessity that is substantive in the current sense, in that a person might well grasp it and entertain it without seeing that it is true, even though it is not Mackie-substantive, in that the error theory does not entail that it is false.

is a conceptual truth in the wide sense just in case (a) the truth of $\langle \text{Necessarily } p \rangle$, and therefore the necessity of p , is entailed by intrinsic properties that p has – or that are had by propositions that entail p as well as by the proposition that p is so entailed – and (b) these are properties that a person must grasp, and grasp that the relevant propositions have, in order to grasp these propositions. For any such proposition p , clause (a) would explain its necessity and clause (b) would explain its a priori epistemology. It follows that if a proposition is a conceptual truth, then a person who met the minimal conditions required for grasping it (or for grasping propositions that entail it as well as the proposition that it is so entailed) would thereby be in a position to know, merely by entertaining it (or it as well as these other propositions), that it is true.

The epistemology of conceptual truths is not my concern. I am interested here in the modal status of substantive moral necessities, which is a metaphysical rather than an epistemological issue. However, if there is reason to believe that some substantive moral necessities have the epistemological characteristics of conceptual truths, then there may be reason to believe that they can be given the kind of metaphysical explanation provided in clause (a) of the above account. For this reason, I need to consider the epistemological issues.

Let us again consider the painful death proposition. In this section I will focus on epistemological issues raised by the idea that it is a conceptual truth, turning to metaphysical issues in the [next section](#). The key metaphysical point is that the proposition appears not to have internal truth conditions, for the wrongness property appears not to be a constituent of the property an action can have of causing a child to die a prolonged and excruciatingly painful death. I will address this point in the [next section](#).

If the painful death proposition is a strict conceptual truth, then we could not grasp it and entertain it – considering carefully whether it is true, and thinking clearly – without being in a position to know it is true. If it is a *derivative* conceptual truth, then it is entailed by other propositions that are conceptual truths and it is a conceptual truth that it is so entailed. Perhaps, for instance, it is entailed by a moral principle that is a conceptual truth. Given the complexities, I obviously cannot hope to prove that the painful death proposition is not a conceptual truth. But I shall offer three arguments. The arguments admittedly are not decisive.

First, consider the viewpoint some slave-holders appear to have held in the early nineteenth century in the United States, a viewpoint according

to which slaves are simply property and can rightfully be treated by the owner as the owner sees fit.²⁶ On this view, there may not be any moral reason (of the kind that tends to make actions wrong) for a slave-holder not to cause a child slave of his to undergo a painful death. To do this would be wasteful and perverse and callous, on the slave-holders' view, but it would not be wrong, not even *other things being equal*. The view is that slaves are living property just as draft animals are living property, and the morality of the treatment of slaves by their owners is a special case of the highly permissive morality of the treatment of animals by their owners. It holds that the moral standards that govern the treatment of free children are completely different from the standards that govern the treatment of child slaves by their owners.²⁷

The example undermines the claim that the painful death proposition has intrinsic truth-ensuring properties, properties that the slave-holders are in a position to grasp in virtue of meeting the minimal conditions necessary for grasping the proposition. For if it has such properties, then something must explain the slave-holders' failure to see that the proposition is true. They surely meet the minimal conditions necessary for grasping it. They have moral views about the treatment of slaves and they disagree with us about the painful death proposition. If so, they must grasp the proposition. Moreover, it is not plausible that the problem is simply the slave-holders' lack of attention to the content of their beliefs or lack of logical acumen. It is not that they simply fail to work out that the proposition is true.

One might think that the problem lies with the slave-holders' overall view, which must be conceptually false.²⁸ But this seems to be a mistake. Their view is beyond the pale, but it is coherent. To be sure, it may well be that the views of at least some of the slave-holders are not in a stable wide reflective equilibrium.²⁹ The slave-holders must view ownership as of enormous moral significance, and this may be something many of them

26 The novel *The Known World* presents a persuasive picture of the psychological and moral complexities of the slave society of antebellum Virginia (Jones 2003). Some characters in the novel appear to accept the perverse moral view I sketch in the text.

27 Similarly, many people think that the moral standards regarding the treatment of children are completely different from the moral standards regarding the treatment of nonhuman animals.

28 A proposition is conceptually false just in case it is the negation of a conceptual truth. There are propositions that are strict conceptual falsities as well as propositions that are derivative conceptual falsities.

29 The notion of reflective equilibrium comes from Rawls 1971.

could not accept on reflection. Moreover, it may be that few slave-holders would agree that if their own child were a slave, it would be permissible for the child's owner to cause the child a painful death. Many slave-holders may be committed by their moral view to the truth of propositions that they could not accept on reflection. The important point, however, is that the view appears to be coherent.

The example undercuts the idea that the painful death proposition is a strict conceptual truth, and it also undercuts the idea that it is entailed by a general moral principle that is a strict conceptual truth. For if it were so entailed, then that principle would be incompatible with the slave-holder's moral viewpoint. It would follow that the slave-holders would be in a position to know that their moral viewpoint is false merely in virtue of meeting the minimal conditions necessary for grasping and entertaining the general principle in question, given that they already grasp and can entertain their own moral view. It seems implausible on its face that this is so.

The second argument is a close relative of G. E. Moore's open question argument. I will present it in general terms. The painful death proposition is of the form $\langle \text{Any action that is } P \text{ has the wrongness property} \rangle$. For now, let us ignore the difference between the wrongness property (or wrongness other things being equal), and wrongness *simpliciter*. There are two cases to consider. First, suppose that the property an action could have of being P is partly constituted by wrongness such that an action's being P entails that it is wrong, just as an action's being a schmurder entails that it is wrong. In such cases, as we have seen, the proposition in question is not Mackie-substantive. Second, however, suppose that the property of being P is not partly constituted by wrongness. In this case, it seems, it is not a conceptual truth that an action's being P entails that it is wrong. Hence, it seems, a person could grasp and entertain the proposition that $\langle \text{Any action that is } P \text{ has the wrongness property} \rangle$ while doubting that it is true. As Moore might insist, the fact that, on the current supposition, we understand what would be involved in doubting the proposition is at least *prima facie* reason to think that it is not a conceptual truth.³⁰ To be sure, some conceptual truths are not *obviously* true. But the burden of argument

30 Moore says, "It may indeed be true that what we desire to desire is always also good; . . . but it is very doubtful whether this is the case, and the mere fact that we understand very well what is meant by doubting it, shews clearly that we have two different *notions* before our minds." Moore 1903, sec. 13, p. 68, emphasis added.

now falls on my opponent to show that the P proposition is a substantive truth that is also a conceptual truth.

The final argument turns on the idea of a Mackie-substantive moral proposition. Mackie met the minimal conditions required for grasping the painful death proposition, yet he would have denied that the proposition is true on the ground that, he thought, nothing could be wrong. That nothing can be wrong is entailed by his error theory. Now I believe that the error theory is false, but I deny that it is conceptually false. One might think that some of Mackie's arguments for the theory rested on incoherent claims. Even if so, it does not follow that the theory is conceptually false. Indeed, I think it plainly is *not* conceptually false.³¹ It is not a conceptual truth that there is a property of wrongness – or that there is 'such a thing' as wrongness. At the very least, if it is a derivative conceptual truth that the error theory is false, the argument that it is false is not a simple one-step argument in the way that it would be if the painful death proposition were a strict conceptual truth. But then, since the painful death proposition entails that the error theory is false, it follows that this proposition is not a strict conceptual truth. For if it were, then the error theory would be (derivatively) conceptually false, and by a simple and obvious argument.

This argument also undercuts the idea that the painful death proposition is entailed by a general moral principle that is a conceptual truth. For if it were so entailed, then that principle would be incompatible with the error theory and could be used to argue directly that the error theory is conceptually false. But it is implausible that this is so. Hence, it is implausible that any general principle that entails the painful death proposition is a conceptual truth. The argument can be generalized. Mackie's error theory entails the falsity of *any* substantive moral proposition. Hence, by *modus tollens*, any such proposition entails the falsity of the error theory. If some such proposition were a *conceptual* truth, then the error theory itself would be derivatively conceptually false, and by a simple argument. But this is implausible. So it is implausible that there is a substantive moral proposition that is a conceptual truth. That is, plausibly, any moral proposition that is a conceptual truth is not Mackie-substantive.

One might object that, despite what I have argued, the painful death proposition seems to be *self-evident*. I can agree with this, for the idea of self-evidence is best understood in such a way that a self-evident

31 For arguments that it is not conceptually false, see Sinnott-Armstrong 2006, ch. 3.

proposition need not be a conceptual truth. According to Russ Shafer-Landau, “A proposition p is self-evident =_{df} p is such that adequately understanding and attentively considering just p is sufficient to justify believing that p .”³² Shafer-Landau does not explain what he takes to be involved in ‘adequately understanding’ a proposition, but he seems to have in mind a richer appreciation of what would be involved in the truth of a proposition than what is required merely to grasp it.³³ Hence, on his definition, a proposition might be self-evident even if one would not be in a position to know it to be true merely in virtue of meeting the minimal conditions necessary for grasping it.³⁴ Moreover, the property a proposition can have of being self-evident is epistemological. Even if the painful death proposition is self-evident, this has no metaphysical implications, and so it can provide no explanation of the proposition’s necessity.

To conclude this section, *substantive* moral necessities seem not to be conceptual truths – neither strictly speaking nor derivatively. Although the arguments are not conclusive, I think they put the burden of proof on anyone who thinks otherwise.

4. SUBSTANTIVE MORAL NECESSITIES AS METAPHYSICALLY NECESSARY

One might wonder whether there are necessary truths that do not qualify as conceptual necessities because they lack the necessary epistemological characteristics. I will speak of such propositions, if there are any, as “metaphysical necessities.” The question we need to ask is whether substantive moral necessities are metaphysical necessities.

A familiar putative example of a metaphysical necessity is the proposition that any quantity of water is a quantity of H_2O .³⁵ Call this “the H_2O proposition.” Its status is controversial. If we assume that the property of being water is identical to the property of being H_2O , then the proposition is necessarily true. On this assumption, however, it may seem that the

32 Shafer-Landau 2003, 247. See also Audi 1997. I discuss the notion that moral propositions may be self-evident in chapters 2 and 3.

33 Shafer-Landau 2003, 279 n. 12. See chapter 3.

34 Shafer-Landau’s definition does not guarantee that a self-evident proposition is true, which is an additional reason why a proposition could be self-evident without being a conceptual truth. Of course, one could amend his definition.

35 This example has been discussed by many philosophers, beginning with Hilary Putnam (Putnam 1975).

proposition is a conceptual truth. For it is arguable that we cannot grasp it without knowing what water *is*, and so, if water is H_2O , we cannot grasp it without being in a position to know that water is H_2O . The issue depends on complex questions about the nature of propositions and about what is involved in grasping a proposition.

The controversy about the H_2O proposition raises the question whether it is possible for a proposition to have intrinsic truth-ensuring properties without being a conceptual truth. The issue is whether one could grasp a proposition that has such properties without being in a position to know it is true simply on the basis of what one must grasp in grasping it. On the one hand, the properties in question are intrinsic to the proposition, and it is arguable that grasping a proposition puts a person in a position to grasp its intrinsic properties. If so, then a person who grasps a proposition with intrinsic truth-ensuring properties is thereby in a position to know that it is true. It would follow that it is a conceptual truth. On the other hand, if any proposition has intrinsic truth-ensuring properties, surely the H_2O proposition does, at least if we assume that the property of being water is identical to the property of being H_2O . And it is intuitively plausible that one can grasp the H_2O proposition without being in a position to know that it is true. Intuitively, in order to grasp the proposition, we need only know something like this: Water is the “watery stuff of our acquaintance,”³⁶ and H_2O is the substance the molecules of which are composed of one oxygen atom bonded with two hydrogen atoms. We do not need to know that H_2O is the chemical composition of water. Hence, it seems, we could grasp the proposition without being in a position to know it is true. The H_2O proposition is therefore not a conceptual truth.³⁷ To explain this, we have to allow that grasping a proposition does not require being in a position to grasp all of its intrinsic

36 I take this felicitous wording from Jackson 1998, 38.

37 On the account of “structured propositions” given in King 1998, the terms “water” and “ H_2O ” could both contribute the property of being water – i.e., the property of being H_2O – to the proposition expressed by the sentence “Water is H_2O ,” even though this proposition is distinct from the proposition expressed by the sentence “Water is water.” King’s idea is that the different terms make distinct contributions to such propositions since, unlike “water,” the use of “ H_2O ” brings some of the constituents of the property of being water into such propositions. King’s view appears to have the power to explain why one who grasps the proposition expressed by the sentence “Water is H_2O ” might not thereby be in a position to know, merely by considering it carefully, that it is true. Hence, it appears to have the power to explain why this proposition is not a conceptual truth even though the property of being water is the property of being H_2O and even though the proposition expressed by the sentence “Water is water” is a conceptual truth.

properties. If this is correct, a proposition with intrinsic truth-ensuring properties need not be a conceptual truth.

Consider, then, the set of propositions that are necessary truths (in the strongest sense). Some of them are strict conceptual truths and some are derivative conceptual truths. The remaining necessary truths, if any, are ‘synthetic.’³⁸ I call them “metaphysical necessities.”

The truth of a necessary truth ‘at a world’ cannot depend on facts about that world that distinguish it from other worlds. Its truth conditions are ‘world-insensitive.’ This seems to mean that the necessity of a metaphysical necessity must be explained either by something about it that guarantees its truth or by something that is invariant among worlds. Perhaps it carries its truth with it from world to world, so to speak. In this case, the proposition must have internal truth conditions or intrinsic truth-ensuring properties or be entailed by propositions with intrinsic truth-ensuring properties. Or perhaps its truth is guaranteed by conditions common to all possible worlds. In this case, the proposition manages to be true in all worlds because the state of affairs it represents as obtaining obtains in all worlds.

On the account I gave in the [previous section](#), the necessity of a conceptual truth also has a metaphysical explanation; it is explained either by the fact that the proposition has internal truth conditions or by the fact that the proposition is entailed by propositions with internal truth conditions. So in considering whether there are substantive moral necessities that are metaphysical necessities, we are also addressing what must be the case if there are substantive moral necessities that are conceptual truths.

With this background, consider again the painful death proposition. The necessity of this proposition depends on there being a relation between two properties such that the instantiation of the first necessitates the instantiation of the second. The first is the ‘painful death property’ – the complex property an action can have of being an intentional causing of a child’s prolonged and painful death – and the second is the ‘wrongness property’ – the property an action can have of being wrong other things being equal. The necessity of the proposition depends on whether the instantiation of the first entails the instantiation of the second.

38 I stipulate that a proposition is ‘synthetic’ just in case it is neither a conceptual truth (in the wide sense) nor the negation of a conceptual truth (in the wide sense). Notice that I am speaking here of propositions as “synthetic.” I used this notion of a synthetic proposition in chapter 1.

Some relations between distinct properties hold of necessity. For example, if one property, 'Nness,' is a constituent of another property, 'Nplus-ness,' this relation holds of necessity and anything that is Nplus obviously must be N. Intuitively, for example, the property of being a bachelor is partly constituted by the property of being unmarried, so there is no possibility of instantiating the former without also instantiating the latter. This explains why it is necessary that a bachelor is unmarried.

Intuitively, however, the wrongness property is not a *constituent* of the painful death property. An action may be wrong other things being equal *because* it has the painful death property. But intuitively, this would be a matter of the action's having the *further* characteristic of being wrong other things being equal, where it has this characteristic *due to* its having the painful death property. Intuitively, it isn't a matter of the wrongness property's being a constituent of the painful death property in the way that being unmarried is a constituent of being a bachelor. Being unmarried is not something further that is true of a bachelor, in addition to his being a bachelor, which is explained by his being a bachelor; something does not qualify as a bachelor unless it is 'already' unmarried. If someone says, awkwardly, that someone is unmarried "because" he is a bachelor, we hear the "because" as epistemic; we hear the person explain her belief that the man is unmarried by citing the fact that he is a bachelor. In the moral case, however, the "because" is not merely epistemic. We seek to explain why an action was wrong, and not merely why we believe it was wrong, by citing the fact that it was an intentional causing of a painful death.

Another relation between distinct properties that holds of necessity, if it holds at all, is the relation between a disjunctive property and one of its disjuncts. Assume that there is a property of being either gold or copper. Call it the property of being "gopper." There is no difficulty understanding why it is necessary that any sample of gold is gopper. There is no possibility of something's being gold without being gopper, since being gold is one of the disjunctive ways of being gopper. In general, if the property 'Ndisjunction-ness' is a disjunctive property, and if another property, 'Nness,' is one of the disjuncts, this relation between the properties holds of necessity, and anything that is N obviously must be Ndisjunction.

Intuitively, however, the wrongness property is not a disjunctive property of this kind. It is not the property of either being thus and so, or being such and such, or being an intentional causing of a painful death, or whatever. Actions that are wrong have something in common. An action may be wrong other things being equal *because* it is an intentional

causing of a child's prolonged and painful death. But if so, as I said, this is a matter of there being something *further* that is true of intentional causings of painful deaths, namely, that they are wrong other things being equal. Their being wrong other things being equal is explained by their being intentional causings of painful deaths. It isn't simply that being an intentional causing of a painful death is a disjunctive way of being wrong other things being equal, as being gold is a way of being gopper. Being gopper is not something further that is true of gold, in addition to its being gold, which is explained by its being gold; it is just a matter of gold's either being gold or copper. Something may be gopper "because" it is gold, but, intuitively, this "because" relation is merely classificatory. In the moral case, however, the crucial relation is explanatory rather than classificatory. We explain why an action was wrong by citing the fact that it was a causing of a painful death.

If the painful death proposition is necessarily true, this is presumably due to the fact that an action that has the painful death property has the wrongness property *because* it has the painful death property. Alternatively, we might say, the fact that an action has the painful death property *makes it the case* that it has the wrongness property or *brings about* that it has the wrongness property. That is, the 'bringing about' relation obtains of necessity between the painful death property and the wrongness property. We can ask what explains this, but this is simply to move our problem back one step. We began with the problem of explaining the necessity of the painful death proposition. We now have the problem of explaining the necessity of the proposition that the fact that an action has the painful death property *brings about* that it has the wrongness property.

Given what I have said, it is difficult to see how we could explain the necessity of the painful death proposition. It appears to be neither a conceptual truth nor a metaphysical necessity. Its truth appears not to be guaranteed by the intrinsic nature of either the painful death property or the wrongness property. If it is necessarily true, its necessity must be explained in some way I have not yet considered.

5. MORAL NATURALISM AND THE MORAL NECESSITIES

I shall now turn the tables on these arguments. For despite what I have been arguing, certain kinds of moral naturalism are able to explain the necessity of the painful death proposition, and, if they are correct, the proposition is metaphysically necessary. The important thing about these

theories is that they each propose a thesis I will call a “T-identity thesis”; a thesis of the form $\langle \text{Nness is identical to Mness} \rangle$ for a moral predicate “M” and a (perhaps very complex) nonmoral predicate “N.” Expressed metalinguistically, the thesis is that the property ascribed by “M” is identical to the property ascribed by “N.”³⁹

Now for any T-identity thesis, there is a corresponding proposition of the form $\langle \text{Anything that is N is M} \rangle$, which I will refer to as the corresponding “T-necessity.” This proposition, the T-necessity, is a substantive moral necessity, assuming that the T-identity thesis is true and that there are possible worlds in which something is N. It is a necessary truth on these assumptions, for if Nness is identical to Mness, then it is not possible for something to be N without being M. It is substantive on these assumptions because, if there are worlds in which something is N, then Mackie’s error theory entails that it is not the case that $\langle \text{Necessarily, anything that is N is M} \rangle$. Hence, theories that propose a T-identity thesis are able to ground at least some substantive moral necessities. The interesting question is whether they can ground intuitive moral necessities, such as the painful death proposition.

I shall focus on forms of moral *naturalism* that propose T-identity statements.⁴⁰ Although nonnaturalistic theories typically hold that moral properties are unanalyzable, a nonnaturalistic theory could in principle propose a T-identity thesis. For example, a divine command view might propose that obligatoriness is identical to the property of having been commanded by God. This could be classified as a nonnaturalistic view, assuming that the property of being commanded by God is not a natural property.⁴¹ Much of what I say would apply, *mutatis mutandis*, to any theory that proposed a T-identity statement.

It will be useful to distinguish between two relevant kinds of naturalism. ‘Conceptual naturalism’ proposes a T-identity thesis and claims it is a conceptual truth; ‘metaphysical naturalism’ proposes a T-identity thesis

39 Nicholas Sturgeon denies that moral naturalism is committed to defending a T-identity thesis (2006).

40 Moral naturalism holds that moral properties are ‘natural properties.’ Hence a naturalistic T-identity thesis identifies Mness with a natural property Nness. In chapter 1, I propose that a property is a ‘natural property’ just in case, roughly, any synthetic proposition about its instantiation that can be known can only be known empirically.

41 One might object that ‘nonnaturalism’ is commonly taken to be antireductionist and that, as such, it does not propose a T-identity thesis. One might even take this to be true by stipulation. The view mentioned in the text could then be described as ‘not-naturalist.’

and claims that it is metaphysically necessary but does not claim that it is a conceptual truth. The difference between these two kinds of naturalism is epistemological. Conceptual naturalism combines the metaphysical claim that its T-identity thesis is true with the epistemological claim that this thesis can be known merely in virtue of grasping and entertaining propositions with internal truth conditions. The latter claim is not important for our purposes.

It might be thought that conceptual naturalism has been defeated by Moore's open question argument.⁴² In capsule form, Moore's argument is that, for any T-identity thesis, it is possible for a person to grasp it and entertain it yet to doubt whether it is true; given this, Moore holds, the thesis must be false, or at least, it cannot be a conceptual truth. A defender of conceptual naturalism has an effective response, however. For even a strict conceptual truth may not be obviously true, and if a T-identity thesis is a derivative conceptual truth, then even though it is knowable simply on the basis of grasping and entertaining propositions with truth-ensuring properties and working out their entailments, it may be a difficult matter to work out the entailments. It is therefore possible that a T-identity thesis is a conceptual truth even if it appears to be an open question whether it is true.

To illustrate differences among T-identity doctrines, especially differences in their ability to explain substantive moral necessities, I will briefly discuss four examples. I begin with an example of conceptual naturalism and then turn to three examples of metaphysical naturalism. The point here is to look at what they can say about moral necessities.

Michael Smith defends a T-identity thesis, which I believe he takes to be a derivative conceptual truth. Setting aside details that are not important for my purposes, he holds that the property of being morally required is the property 'Fness' such that any ideally rational person who was fully informed would want us to perform actions that are 'F,' where Fness is a property of an appropriate kind.⁴³ This thesis grounds the necessity of the corresponding T-necessity, the proposition that any action that is F is morally required, where Fness meets the conditions specified in the T-identity thesis. Unfortunately, Smith's theory cannot explain the necessity of the painful death proposition. At least, it cannot do this unless it can support the necessity of the proposition that any ideally rational person who was fully informed would want us (at least other things being

42 Moore 1903, sec. 13, pp. 66–68.

43 Smith 1994, 28–35, 182–187.

equal) not to cause painful deaths. I do not see any machinery in the theory that would entail the necessity of this proposition or of the similar propositions that would have to be necessary truths in order for the theory to underwrite other intuitive moral necessities.

The first kind of metaphysical naturalism that I will consider is a kind of theory that invokes a general semantic view about the relation between predicates and properties to argue that moral predicates ascribe natural properties. Richard Boyd has proposed a theory of this kind, invoking a causal-explanatory semantics.⁴⁴ One could instead invoke a Putnam-style semantics that turns on facts about speakers' semantic intentions.⁴⁵ Positions of this kind are contested,⁴⁶ but we can see how they would support an account with the potential to explain the truth of certain substantive moral necessities. Boyd aims to defend, for example, the T-identity thesis that moral goodness is identical to a certain natural property, the "homeostatic cluster property" that, he argues, is ascribed by the term "morally good" in standard contexts. On this account, it turns out to be a T-necessity that anything that is morally good has this property. Boyd is a consequentialist, and if he is successful in defending the necessity of the consequentialist principle he favors, the necessity of the painful death proposition may follow.

The basic strategy here must be correct, it seems to me. The semantics of moral predicates needs to be integrated smoothly into a general semantics for predicates. The trouble is that we do not have a general theory of the semantics of predicates that is sufficiently well-developed to enable us to determine which natural property, if any, is ascribed by a given moral predicate. Because of this, we cannot be confident that Boyd's proposal is correct, and we cannot be confident that a theory of this kind will explain the necessity of the painful death proposition.⁴⁷

The second kind of metaphysical naturalism that I will consider is the "moral functionalism" that has been proposed by Frank Jackson and Philip Pettit.⁴⁸ There are three central ideas. First is the idea that the moral facts supervene on the descriptive facts, which, they argue, entails that moral properties are natural or 'descriptive' properties.⁴⁹ Second is

44 Boyd 1988.

45 See chapters 6 and 7.

46 See, e.g., Horgan and Timmons 1992.

47 In chapter 7, I explain how one could marry a Putnamian semantic theory to my own society-centered proposal.

48 Jackson and Pettit 1995; Jackson 1998, 129–151.

49 Jackson 1998, 122–123.

the idea that there is or may be a moral theory – they call it “mature folk morality” – on which current folk morality will converge in the long run, given appropriate reflection and full empirical information.⁵⁰ And third is David Lewis’s strategy for defining theoretical terms, which Jackson and Pettit employ in order to identify the descriptive properties in question.⁵¹ In effect, their proposal is to view mature folk morality as specifying “roles” that the various moral properties play. Presumably, for example, there will be a ‘wrongness role,’ which will be understood partly in terms of its relation to other roles specified by mature folk morality, such as the ‘injustice role.’ These relations will in turn be specified partly in terms of ‘commonplaces’ or ‘platitudes’ which are such that the acceptance of them (or at least the tendency to accept them) is a necessary condition of competence in moral discourse and thought.⁵² For example, it presumably is a commonplace that injustice is wrong. The moral properties are to be identified by the relevant roles in mature folk morality. For example, wrongness is identified in terms of the wrongness role. There are two relevant possibilities. Wrongness might be the “realizer property” – “the (first order) descriptive, possibly disjunctive, property that plays the [wrongness] role.” This could be, say, the property of failing to maximize the good. Alternatively, wrongness might be the “role property” – “the second-order property of having the property that plays the rightness role.” Jackson argues that the first option is preferable.⁵³ The upshot is that the identity of the property ascribed by “wrong” will depend on the content of mature folk morality. Wrongness is a natural property, but exactly which natural property is an a posteriori matter.⁵⁴

A theory of this kind could in principle defend the necessity of the painful death proposition, for it might turn out that the proposition is one of the platitudes, and if so, acceptance of it (or a tendency to accept it) is constitutive of competence in moral discourse. It might turn out, on the best account, that wrongness is a disjunctive property and that the painful death property is one of its disjuncts. Or it might turn out, on the best version of the theory, that a form of consequentialism is necessarily

50 Ibid., 140, 151.

51 Ibid., 133–134, 139–141; Lewis 1970.

52 Jackson and Pettit 1995, 22–23.

53 Jackson 1998, 141; Jackson and Pettit 1995, 27–28.

54 Jackson 1998, 150–151.

true and that it entails the necessity of the painful death proposition. All of this is, however, conjectural.

My central objection to Jackson and Pettit's moral functionalism is that it appears to leave no room for the idea that mature folk morality – the moral theory on which we would converge in the long run, given appropriate reflection and full nonmoral information – might be mistaken. A different problem is that people's moral views might not converge even in the long run, and even given appropriate reflection and full nonmoral information.⁵⁵ Perhaps, for example, the slave-holders I discussed before would continue to deny the painful death proposition. But even if our moral views *would* converge in the long run on a single mature folk morality, it seems that this morality could be mistaken. Jackson appears to object that this objection “is a holdover” from an objectionable platonism about ethical properties.⁵⁶ As I will explain, however, my own society-centered view implies that mature folk morality might be mistaken without invoking the kind of platonism that Jackson appears concerned to avoid.

The third kind of metaphysical naturalism I will discuss is my own ‘society-centered’ theory.⁵⁷ The basic intuition is that morality has the function of making society possible by laying down rules governing our lives that, when they have currency in society, enable people living in the society to get along together better than they otherwise could. Different moral codes differ in how well they would serve this function. The truth about what morality requires depends on what is required by the moral code, the currency of which in society would best serve this function. Two basic ideas underlie this approach: the ‘standard-based’ theory of normative propositions and the society-centered theory of morality.

The standard-based theory proposes a schema for giving the truth conditions of moral propositions in terms of a relevant status of relevantly corresponding standards. It rests on a distinction between moral *propositions* – such as the painful death proposition – and moral *standards* or *norms* – such as the norm that would be expressed by the imperative, “Do not cause a child to die a prolonged and excruciatingly painful death!” As I

55 For discussion of what moral functionalism would be committed to in this case, see Jackson 1998, 137. See also Sinnott-Armstrong 2006, 55.

56 Jackson 1998, 151.

57 Copp 1995.

use the term, standards are not objects of belief. Most of us *subscribe* to the standard that would be expressed by the above imperative, in that, among other things, we are inhibited from causing painful deaths and would feel guilty if we did, but it makes no sense to suppose that someone believes it. Applied to the painful death proposition, the standard-based theory says, roughly, that the proposition is true just in case the corresponding standard, the aforementioned rule that prohibits causing painful deaths, has a relevantly authoritative status. The key idea is that there is *some* status such that when a norm enjoys that status, the fact that it does underwrites the truth of corresponding moral propositions. Call this the “truth-grounding status.”⁵⁸

Society-centered theory provides a theory about the status that a moral norm must have in order that corresponding moral propositions be true. It links the truth conditions of moral propositions to the status that corresponding standards have when their serving in the relevant society as the societal moral code would enable the society better to serve its needs than would the currency of alternative sets of rules. What is a *society*? In brief, and roughly, a society is a multi-generational temporally extended population that is characterized by a network of social relationships that permit cooperation in pursuit of the necessities of life and the priorities of the group’s culture. I explain this in more detail elsewhere.⁵⁹ Which is the *relevant* society? I address this question shortly. The moral implications of the theory are contingent on how best to promote societal needs, which is clearly an empirical matter.⁶⁰ Societies have the same basic needs, but they can be in different circumstances; hence, the moral codes justified in relation to them are unlikely to be exactly the same. This point gives rise to worries of various kinds, but I have to set most of them aside.⁶¹ In our circumstances, I think the upshot is likely to be a familiar kind

58 More generally, the standard-based theory says that a (pure, basic) normative proposition of type K is true just in case a corresponding standard of type K has the K-relevant truth-grounding status. A ‘pure’ normative proposition of type K has no non-K-normative entailments or presuppositions (other than those given by the norm-based theory itself). The proposition that Smith was wrong to lie to Jones is impure. A ‘basic’ normative proposition is subject-predicate in form and (otherwise) logically simple. The proposition that either lying is wrong or the moon is full tonight is nonbasic.

59 Copp 1995, ch. 7. See also the [introduction](#) to this book.

60 I am ignoring certain complications that are not relevant to the issues being addressed in this chapter. I explain the idea of a basic need in Copp 1995, ch. 9. I discuss the basic needs of societies at pp. 192–194. See also the introduction to this book.

61 See Copp 1995, 201–209.

of deontological moral code. Such a code would impose familiar duties on us, including, I believe, the duty not to intentionally cause a child to undergo a prolonged and excruciatingly painful death.

Society-centered theory supports the T-identity thesis that the predicate “wrong (all things considered),” when used in a given context, ascribes, very roughly, the property of being ruled out by the moral code that is actually the code the currency of which in society S would better enable S to meet its needs than would the currency of any alternative, where S is the society that is relevant in the context. If moral code M is actually this code, then the theory says that the property that is in question when we talk of an action as being “wrong (other things being equal)” is the property, roughly, of being ruled out by a norm that is entailed by or included in M. On this approach, moral properties are best viewed as relational. Wrongness is instantiated by an action when the action stands in the relation I have just defined to a moral code and a relevant society.

Which is the relevant society? In the default case, it seems to me, the society at issue in a given context is the smallest society that embraces all the people referred to in the context, all the people the speakers or thinkers have in mind in the context, and all the people who are party to the conversation.⁶² So, for example, if an American and a Frenchman are debating whether capital punishment is morally acceptable, the relevant society is, presumably, the very large society that embraces the populations of all Western democracies. If a philosopher is considering whether to accept consequentialism, the relevant society is likely to be the society of all rational persons, since it is likely that the philosopher views morality as prescribing duties for all rational persons. The default does not always obtain. If we are wondering whether something would be acceptable in a society that is in radically different circumstances from our own society, the relevant society is the other society, not our own society.

The theory says that different propositions may be expressed in different contexts by the “painful death sentence” – the sentence “It would be wrong intentionally to cause a child a prolonged and excruciatingly painful death” – depending on which society is relevant in the context (even if we restrict attention to contexts in which the words retain their standard meaning in English). Let us say that in the current context, ‘our society’

62 See *ibid.*, 221–223.

is the relevant one. Let us assume that moral code ‘M,’ the currency of which would best serve our society’s needs, as matters actually are, would entail a prohibition on causing children painful deaths, a prohibition that could be expressed by the imperative, “Do not cause a child to undergo a prolonged and excruciatingly painful death!” On this assumption about the content of M, the proposition that would be expressed in the current context by the painful death sentence is true. For M entails a norm that prohibits causing a child to undergo a painful death.

There are possible contexts, however, in which an assertion of the painful death sentence would express a proposition different from the one it would express if asserted in our current context (even if the words retain their standard meaning in English). And this different proposition might *not* be true. I have in mind contexts in which the relevant society faces circumstances such that a rule that prohibits causing painful deaths is *not* entailed by the moral code the currency of which in the society would better enable it to meet its needs than would the currency of any other moral code. Perhaps the society faces such difficult circumstances that it is not in its interests for the societal moral code to entail a prohibition of causing painful deaths. Suppose an evil genius will destroy the society if the moral code entails a prohibition of causing painful deaths. In this case, causing painful deaths would not be wrong (other things being equal) relative to that society, according to the theory.

This implication of the theory ought to seem counter-intuitive if the theory is correct. For if we have internalized moral code M, the code that is the best code for our actual circumstances according to the theory, or if we have internalized a code that closely approximates to M, then on my assumption about the content of M, we will think that it would be wrong (other things being equal) to cause a painful death notwithstanding the threat of the evil genius.⁶³ The fact that the theory predicts in this

63 Let me explain more fully. The theory predicts that the moral code each of us is taught during our socialization will tend to consist of rather simple rules, such as, I am supposing, an unqualified prohibition on harming children. There is no gain to a society in having the moral rules be complex enough to anticipate every possible eventuality, including the possible threat of an evil genius, given how improbable this is. And it would be counter-productive to have the rules be this complex, since, even if such complex rules could be learned and internalized, we do not want people to have the idea that, for example, it may be permissible in some circumstances to harm a child. For reasons of this kind, I think it is likely that moral code M will include a simple unqualified prohibition on harming children rather than a prohibition that is qualified by, for example, a permission to harm children if it happens that an evil genius is threatening to destroy society unless a child is harmed. See Copp 1998.

way that some of its normative implications will seem counter-intuitive means that the fact that it has counter-intuitive implications of this kind is not straightforwardly an objection to it.

The important point for present purposes is that society-centered theory says there are contexts in which the painful death sentence does not express a necessary truth (even if the words retain their standard meaning in English). These are not contexts, however, in which this sentence expresses ‘the’ painful death proposition. The issue we have been discussing, is whether and how we can ground the necessity of ‘the’ painful death proposition. This proposition is, presumably, the proposition *we* actually express by means of the painful death sentence. According to society-centered theory, then, the issue should be framed as the question whether the proposition expressed by the painful death sentence in the *actual* context is a necessary truth, and if so, how we can ground its necessity.

I have assumed that M – the moral code the currency of which would best serve our society’s needs as matters actually are – would entail a prohibition on causing painful deaths, a prohibition that could be expressed by the imperative, “Do not cause a child to undergo a prolonged and excruciatingly painful death!” Now consider any (morally relevant) possible circumstance. Ask whether it would be wrong other things being equal in that circumstance to cause a painful death. The relevant society is our society, given that the context in which we are asking the question is the actual context, and the relevant moral code is M. We are asking, in effect, whether it would be wrong-relative-to-our-society-and-relative-to-M, in that circumstance – other things being equal – to cause a painful death. We have assumed that M entails a prohibition on causing painful deaths. Hence, according to the theory, it would be wrong-relative-to-our-society-and-relative-to-M, in that circumstance – other things being equal – to cause a painful death. Therefore, the (relevant) painful death proposition – that is, the proposition expressed by the painful death sentence in our actual context – is necessarily true.⁶⁴

It might help if I present the view in general terms – ignoring, for simplicity, the ‘other things being equal’ qualification. According to the

64 A complication is that if we are considering remote actual societies and what would be wrong in circumstances arising for people in those societies, then the theory says that the society relevant to the context is the largest society that embraces all of us as well as all of those people. This could be the global society. Nevertheless, it seems likely that the best moral code for any actual society would include an unconditional prohibition on causing painful deaths. See Copp 1995, 221–223.

society-centered theory, the proposition that would be expressed in a given context by a sentence of the form “Actions that are W are wrong” is a proposition of the form $\langle \text{Actions that are W are wrong-in-relation-to-S/M} \rangle$, where S is the society that is relevant to the actual context, and where M is the moral code the currency of which in S actually would better enable S to meet its needs than would the currency of any other moral code. Moreover, according to the theory, a moral proposition of the form $\langle \text{Actions that are W are wrong-in-relation-to-S/M} \rangle$ is true just in case, where M is the moral code the currency of which in S actually would better enable S to meet its needs than would the currency of any other moral code, M rules out actions that are W. A proposition of this form is a *necessary* truth just in case M rules out actions that are W in any morally relevant possible world. Hence, the proposition that would be expressed in an actual context by a sentence of the form “Actions that are W are wrong” is a necessary truth just in case, where S is the society that is relevant to the context, and where M is the moral code the currency of which in S actually would better enable S to meet its needs than would the currency of any other moral code, M rules out actions that are W in any relevant possible world.

The upshot of this discussion is that certain kinds of moral naturalism can explain the existence of at least some substantive metaphysical moral necessities. Theories of this kind propose T-identity statements of the form, $\langle \text{Nness is identical to Mness} \rangle$, for moral predicate “M” and non-moral predicate “N.” They therefore underwrite moral necessities of the form, $\langle \text{Anything that is N is M} \rangle$. In addition, some theories of this kind would enable us in principle to explain the necessity of the painful death proposition. The society-centered theory is an example. It supports the thesis that the proposition that would be expressed in the current context by the painful death sentence – ‘the’ painful death proposition – is a metaphysical necessity.

Before I conclude, I need to consider other accounts of the moral necessities, even if only briefly. Otherwise, it may seem that there are ways of explaining moral necessities without embracing a T-identity thesis. I will consider two views.

6. PRACTICAL NECESSITIES: A KANTIAN APPROACH

The interpretation of Kant’s moral philosophy is controversial. Contemporary Kantians offer different interpretations of the Kantian view as well as different approaches to extending or modifying it in order to answer

perceived problems. There are approaches that see the Kantian project as proposing a kind of moral realism,⁶⁵ approaches that see the project as compatible with noncognitivism, and approaches that see the project as offering a kind of antirealist cognitivism. Obviously, I cannot explore these alternatives in any detail. What interests me is the viability of a kind of Kantian argument that aims to defend the ‘practical necessity’ of the fundamental Kantian moral principle. Unfortunately, I can only explore this approach here in schematic terms. The point is to assess, even if only in a preliminary way, its ability to explain the nature and grounding of moral necessities.

Elsewhere, I have called the style of argument I have in mind the “self-conception strategy.”⁶⁶ The idea is that there is a way of thinking of oneself such that a rational person who is thinking clearly *must* think of herself this way, but if she does not accept the fundamental Kantian moral principle, she cannot *coherently* think of herself in this way. There are different versions of this kind of argument. It might be said, for example, that a person who does not accept the Kantian principle cannot see herself as *autonomous*, or that she cannot see herself or value herself as a rationally reflective *agent*, acting for *reasons*, or that she is committed to a *practical solipsism*, or that she cannot coherently expect *other* people to respond to the reasons she addresses to them, or the like.⁶⁷ On the self-conception strategy, rational people who are thinking clearly and coherently accept the Kantian principle because they must accept it in order coherently to conceive of themselves in a given way. Since such people must accept it, it is a necessary truth ‘for all practical purposes.’ It is a ‘practical necessity.’

Such arguments do not show that the Kantian principle is a necessary truth strictly speaking. They do not even show that it is true. They show at best that we must accept it. They might not even show this, for they might only show that we must accept the principle *if* we view ourselves in a certain way, leaving it open whether we *must* view ourselves in this way. And even if we must conceive of ourselves in the relevant way, it does not follow that this way of conceiving of ourselves is accurate.

65 Elsewhere I have proposed that Christine Korsgaard’s neo-Kantian view can be interpreted as a kind of realist moral naturalism. On a naturalist reading, Korsgaard’s view would be able to explain the metaphysical necessity of certain moral necessities. She does not intend the naturalist reading however. See chapter 8 in this book and Korsgaard 1996.

66 See chapter 8.

67 I sketch these ideas with apologies, respectively, to the following philosophers: Kant 1785; Korsgaard 1996; Nagel 1970; Darwall 2006.

Perhaps, for example, I must see myself as autonomous, but it is a vexed issue whether my being autonomous is compatible with the truth of determinism. If it is not, and if determinism is true, then I am not autonomous even if I must think of myself as autonomous. Perhaps I must think of myself as autonomous while I am acting, for I must think of myself as acting for reasons, and perhaps I cannot think of myself as acting for reasons unless I think of myself as autonomous. This will not show that determinism is false or that it is compatible with my being autonomous, so for all that the argument shows, it may be that there is a way of thinking of myself that is inaccurate even though I have no alternative but to think of myself in this way. It appears, therefore, that self-conception arguments show at best that we must accept the fundamental Kantian principle. But it does not follow from this that it is true, much less that it is necessarily true. So we do not have in these arguments a way of securing or explaining a moral necessity strictly speaking.

The idea, however, is that if we have *no alternative* but to conceive of ourselves in a certain way, and if we cannot coherently conceive of ourselves in that way unless we accept the fundamental Kantian principle, then the principle is a 'practical necessity.' If the Kantian principle is a practical necessity, perhaps this is all that we need and all that we have any reason to want.

Unfortunately, however, we surely can believe the Kantian principle to be false. On the self-conception strategy, of course, we cannot *coherently* believe it to be false. So the idea must be that Mackie, who is committed to denying all substantive moral claims, cannot coherently deny the Kantian principle. Yet he can deny it. The self-conception strategy finds an incoherence in the conjunction of his denial of the principle with a view he must take of himself, yet it does not find an incoherence in his denial of the principle as such, nor in his moral theory as such, nor in the conjunction of his theory with the denial of the principle. So it appears that the self-conception strategy leaves open the following view: There are certain beliefs about ourselves that we cannot avoid having and that are incompatible with believing the Kantian principle to be false, yet the principle *is* false, and the relevant unavoidable beliefs about ourselves are also false. On this view, the beliefs about ourselves that are supposedly incompatible with denying the Kantian principle are a kind of cognitive *affliction*. The self-conception strategy argues that these beliefs cannot coherently be combined with denying the Kantian principle. The view we are considering does not deny this, but it says that what follows is,

paradoxically, that the goal of having a *coherent* overall system of belief can be in tension with the goal of avoiding *false beliefs*. If this position is left open by the self-conception strategy, as it seems to be, then even if the Kantian principle is a ‘practical necessity,’ this surely does not give it the kind of status we think that moral necessities have.

7. UNCONDITIONAL ATTITUDES: A NONCOGNITIVIST STRATEGY

The second alternative to naturalistic moral realism that I want to explore is a sophisticated noncognitivist strategy. If one thinks of noncognitivism as entailing that moral judgments do not express propositions with a truth value, one might think that it could not accommodate the intuition that the painful death proposition is true, never mind the intuition that it is necessarily true. Some recently proposed versions of noncognitivism, however, such as the “quasi-realist” view that has been defended by Simon Blackburn, aim to accommodate the “realist surface” of moral thought and discourse. Blackburn adopts a deflationist account of the use of the predicate “true” according to which, to call something “true” is simply to affirm it.⁶⁸ On this view, to say that the painful death proposition is true is simply to affirm that it would be wrong to cause a child a painful death. The theory has an account of what state of mind a person expresses in affirming such a thing. The question is what a theory of this kind – a “sophisticated noncognitivism” – should say about the intuition that the painful death proposition is *necessarily* true.⁶⁹

Simpler forms of noncognitivism are characterized by two doctrines, a positive doctrine and a negative one. Sophisticated noncognitivism adds a third doctrine, a defensive doctrine that is intended to accommodate what Blackburn calls the “realist surface.” The positive doctrine is an expressivist claim to the effect that moral utterances in standard contexts express a conative attitude of one kind or another. Blackburn describes these attitudes as “stances,” such as stances of moral approval and disapproval.⁷⁰ The negative doctrine is more vexed, but equally central. A simple version

68 For an overview of the position, see Blackburn 2006. The phrase “realist surface” comes from this essay. See also Blackburn 1988 and 1993.

69 Jonathan Bennett argues that a noncognitivist can view moral judgments, such as the painful death proposition, as necessary without being committed to their being necessarily *true*. Bennett 1993, 462. See below, note 71.

70 Blackburn 2006.

would have it that the utterance of a simple basic moral sentence in a standard context does not express a belief with a truth value. Sophisticated noncognitivism cannot say this, however, for it needs to make room for the defensive doctrine according to which it is not incorrect to speak of ourselves as having ‘moral beliefs,’ or to speak of a moral belief as ‘true,’ or even, perhaps, to speak of wrongness as a ‘property.’ For this reason, a sophisticated noncognitivism needs to formulate its negative doctrine in a much more nuanced and hedged way, along the following lines: A moral utterance in a standard context does not express a state of mind that represents the world as containing a state of affairs involving moral properties – not, that is, properties that would be postulated in an acceptable metaphysics and treated in an acceptable philosophy of language as being ascribed by moral predicates. This is to say, in effect, that a moral utterance does not express a moral belief the content of which is a *moral* proposition. The defensive claim, which is characteristic of sophisticated noncognitivism, says that there is nothing defective about the “realist surface” of moral thought and discourse. We talk about certain moral claims being true and we talk about our moral beliefs. We talk about properties such as wrongness. The claim is that all of these ways of speaking are acceptable.

Given the negative claim, sophisticated noncognitivism would deny that there are ‘metaphysically robust’ moral necessities. That is, it would deny that an acceptable metaphysics would postulate the existence of metaphysically necessary moral propositions that are akin, for example, to the necessity that whales are mammals. This means that sophisticated noncognitivism satisfies the moral necessities constraint by default. It does not postulate the existence of moral necessities, so it does not need to explain what grounds them. Of course, given its defensive claim, it would insist that there is nothing amiss in the intuition that it is necessarily the case that it would be wrong to cause painful deaths. And given its positive claim, it would need to explain what state of mind a person expresses in saying that this is so.

There are different ways to go at this point. No matter what kind of explanation is offered, however, it will not provide a substantive and metaphysically robust explanation of the truth of moral necessities. To be sure, given deflationism about “true,” sophisticated noncognitivism allows that, where “S” expresses a moral claim, “It is necessary that S” is true just in case it is necessary that S. But this is not a metaphysically robust explanation of what might make it the case that it is necessary that S, nor is

it intended to be. For, as we saw, sophisticated noncognitivism would deny that there are metaphysically robust moral necessities. It would hold that there is nothing that would be acknowledged in a plausible metaphysics as constituting the necessity that it is wrong to cause painful deaths. It is simply the case that there is a kind of unconditional conative attitude that a person can properly express by uttering sentences such as “It is necessarily the case that it would be wrong to cause a child a painful death.”⁷¹

It is part of my moral view, at least, that the moral necessities have a ‘robust status,’ a status that is akin to the status of, say, the necessity that whales are mammals, and a status that is not simply a matter of our attitudes. This points to a potential difficulty for sophisticated noncognitivism. Is its defensive claim coherent with its negative claim? Can it sustain the “realist surface” of moral thought and discourse while also sustaining its negative claim? Trouble will emerge if the realist surface includes intuitions that conflict with the negative claim. If we have the intuition that, necessarily, it is wrong to cause a child to undergo a painful death, and that this necessity has the above robust status, then we have intuitions that conflict directly with the negative claim as I understand it. And since the negative claim is at the heart of noncognitivism, the noncognitivist would have to reject the intuition. He would have to admit that there are exceptions to the defensive claim and that not every aspect of the “realist surface” can be accommodated. If this is correct, then it appears that sophisticated noncognitivism would be unable to do justice to the intuition that there are robust moral necessities. This conclusion must be tentative, however, for sophisticated noncognitivism is still under construction.

8. CONCLUSION

Let me summarize the argument. If we believe that there are robust moral necessities, then we must aim to defend a kind of moral realism. Moreover, if we accept the moral necessities constraint – the constraint that says an adequate metaethical theory would explain the nature and basis of any moral necessity that it postulates – we have reason to opt for a version

71 Bennett proposes that a person who utters the painful death sentence in a standard context would express an unconditioned negative attitude toward anyone’s causing a painful death no matter what the circumstances, and, in this sense, she would accept the judgment as necessary (Bennett 1993, 462). See also Gibbard 1990, ch. 8.

of moral naturalism. For, if I am correct, moral necessities cannot be explained as conceptual or metaphysical necessities unless a theory that proposes a T-identity thesis can be defended. Such theories are typically naturalistic, for it is typical of nonnaturalism to take moral properties to be unanalyzable. The ability of naturalism to account for substantive moral necessities is a good reason to take moral naturalism seriously.

The issue, however, is not merely whether we can explain and defend some moral necessities. The issue is whether we can explain and defend the moral necessities that seem morally plausible, such as the painful death proposition. It is unlikely, of course, that any theory that takes the explanatory task seriously will manage to defend everything that initially seems plausible. The society-centered version of moral naturalism is no exception. Yet it can defend some moral necessities. We have the intuition that ‘the’ painful death proposition – the one we have in mind in thinking that it would be wrong (other things being equal) to cause a child a prolonged and excruciatingly painful death – is necessarily true. I have tried to explain how this intuition can be grounded, given plausible empirical assumptions.

It is important to distinguish the thought that the moral necessities have a kind of metaphysical status that warrants thinking of them as necessary from the thought that they have a kind of epistemological status that sets them apart from other moral propositions. I have argued that the substantive moral necessities that interest us are not conceptual truths, yet there may be other ways to support the thought that they have a special epistemological status. Perhaps, as I mentioned before, they are self-evident. One might think that naturalism is not able to explain their epistemic status.⁷² In other work, I have argued that it can,⁷³ but this has not been my task here. My task has been metaphysical. I have argued that theories that propose T-identity theses are in principle in a position to explain moral necessities that other realist theories must take as unexplained primitives.

REFERENCES

- Audi, Robert. 1997. *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press.
- Bennett, Jonathan. 1993. “The Necessity of Moral Judgments.” *Ethics* 103: 458–472.

72 Fine argues, for instance, that naturalism cannot account for the fact that our moral knowledge is independent of the evidence of our senses. Fine 2002, 272–278.

73 See chapters 2 and 3.

- Blackburn, Simon. 1988. "How to Be an Ethical Antirealist." *Midwest Studies in Philosophy* 12: 361–375.
- . 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- . 2006. "AntiRealist Expressivism and Quasi-Realism." In Copp 2006: 146–162.
- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, N.Y.: Cornell University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1998. "Morality and Society – The True and the Nasty: Reply to Leist." *Analyse und Kritik* 20: 30–45.
- , ed. 2006. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
- Dancy, Jonathan. 2004. *Ethics Without Principles*. Oxford: Oxford University Press.
- Darwall, Stephen. 2006. "Morality and Practical Reason: A Kantian Approach." In Copp 2006: 282–320.
- Fine, Kit. 2002. "The Varieties of Necessity." In *Conceivability and Possibility*, ed. Tamar Szabo Gendler and John Hawthorne, 253–281. Oxford: Clarendon Press.
- Foucault, Michel. 1979. *Discipline and Punish*. Trans. Alan Sheridan. New York: Vintage.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Horgan, Terence, and Mark Timmons. 1992. "Troubles on Moral Twin Earth: Moral Queerness Revived." *Synthese* 92: 221–260.
- Hume, David. 1888. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge. Oxford: Oxford University Press.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Jackson, Frank, and Philip Pettit. 1995. "Moral Functionalism and Moral Motivation." *Philosophical Quarterly* 45: 20–40.
- Jones, Edward P. 2003. *The Known World*. New York: Harper-Collins.
- Jubien, Michael. 2007. "Analyzing Modality." *Oxford Studies in Metaphysics* 3.
- Kagan, Shelly. 1998. *Normative Ethics*. Boulder, Colo.: Westview Press, 1998.
- Kant, Immanuel. 1785. *Grounding of the Metaphysics of Morals*. Trans. James W. Ellington. Rpt., Indianapolis: Hackett, 1981.
- King, Jeffrey C. 1998. "What Is a Philosophical Analysis?" *Philosophical Studies* 90: 155–179.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge: Cambridge University Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Lewis, David. 1970. "How to Define Theoretical Terms." *Journal of Philosophy* 67: 427–446.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt., Cambridge: Cambridge University Press, 1993.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press.

- Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" In *Mind, Language, and Reality*, 215–271. Cambridge: Cambridge University Press.
- Quine, W. V. O. 1961. "Two Dogmas of Empiricism." In *From a Logical Point of View*, 2nd ed., 20–46. Cambridge, Mass.: Harvard University Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. Oxford: Clarendon Press.
- Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. New York: Oxford University Press.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sturgeon, Nicholas L. 2006. "Ethical Naturalism." In Copp 2006: 91–121.
- Thomson, Judith Jarvis. 1990. *The Realm of Rights*. Cambridge, Mass.: Harvard University Press.
- Weatherson, Brian. 2004. "Morality, Fiction, and Possibility." *Philosophers' Imprint* 4: 1–27. <http://www.philosophersimprint.org/004003>.

Part Two

Referring to Moral Properties

Realist-Expressivism: A Neglected Option for Moral Realism

1. INTRODUCTION

Moral realism and *antirealist-expressivism* are of course incompatible positions. They disagree fundamentally about the nature of moral states of mind, about the existence of moral states of affairs and properties, and about the nature and role of moral discourse. The central realist view is that a person who has or expresses a moral thought is thereby in, or thereby expresses, a *cognitive* state of mind. She has or expresses a *belief* that represents a moral state of affairs in a way that might be accurate or inaccurate. The view of antirealist-expressivism is that such a person is in, or expresses, a *conative* state of mind, one that consists in a certain kind of attitude or motivational stance toward something, such as an action or a person. Realism holds that moral thoughts have truth conditions and that in some cases these truth conditions are satisfied so that our moral thoughts are true.¹ Antirealist-expressivism holds, to a first approximation, that the distinctive moral content of a moral thought does not have truth conditions.

Given these contrasts between realism and antirealist-expressivism, the view I shall propose in this chapter might seem surprising, for it combines moral realism with a chief positive doctrine of moral expressivism. I call the

I presented early versions of this chapter to the Social Philosophy and Policy Center conference on moral epistemology, in June 2000, and, in 2001, to the departments of philosophy at Boston University and at the University of Florida. I am grateful to the audiences on these occasions for very useful discussion. I am especially grateful for the helpful comments I received from Kent Bach, Justin d'Arms, Steven Davis, Janice Dowell, Don Hubin, Paul Hurley, Jeffrey C. King, Kirk Ludwig, Steven Rieber, David Sobel, Sigrún Svavarsdóttir, and David Velleman.

1 Cf. Sayre-McCord 1988, 5.

view *realist-expressivism*. It holds that our moral beliefs and judgments represent moral states of affairs and can be accurate or inaccurate to those states of affairs, which is the central realist thesis, but it *also* holds that, in making moral assertions, we express certain characteristic conative attitudes or motivational stances, which is a central positive view of expressivism.

The possibility of a view that combines realism with expressivism in this way has not been widely noticed, despite its many theoretical advantages. One explanation for this is presumably that expressivism is characteristically antirealist, so the idea of combining realism and expressivism might seem untenable. A second explanation is that the differences between realist-expressivism and certain familiar, but distinct, realist and antirealist views are rather subtle. Because of this, it can be easy to confuse realist-expressivism with other views that are quite different. There are at least two possible sources of such confusion. On the one hand, realist-expressivism might be confused with a realist version of a standard kind of *internalism*, which is a doctrine or family of doctrines that I will discuss shortly. But it is not a kind of internalism. On the other hand, realist-expressivism might be confused with a position that combines antirealist-expressivism with *deflationism* about the meaning of “true,” where deflationism is another doctrine or family of doctrines that I will discuss shortly. But realist-expressivism is not any kind of antirealism; it is fully realist, and it does not presuppose deflationism. Let me try to clarify both of these areas of potential confusion here; I will return to them at the end of the chapter, after I have developed my own version of realist-expressivism.

First, it is important not to confuse realist-expressivism with familiar versions of internalism. In the sense I have in mind, internalism is the doctrine that a person who believes or judges she ought to do something must be motivated to some degree to do so.² The internalist idea is that there is a necessary connection, or an internal logical connection – even if it is a defeasible one – between the state of accepting a relevant moral claim and being motivated to act or respond appropriately. Internalism can be combined with moral realism, and the resulting view could be described in terms similar to those I used to describe realist-expressivism. That is, an internalist moral realism could be described as the view that,

2 Stephen Darwall calls this doctrine “judgment internalism” to distinguish it from other internalist doctrines. See Darwall 1983, 54–55. Philosophers sometimes propose weakened versions of internalism by specifying that, for instance, any *rational* person who believed she ought to do something would be relevantly motivated. For an example of this sort of account, see Smith 1994, 61.

first, our moral beliefs represent moral states of affairs and can be accurate or inaccurate to those states of affairs, and second, our moral beliefs entail certain characteristic conative motivational states. Realist-expressivism, however, is not committed to the second half of this view.³ The expressivism in realist-expressivism is basically the thesis that, in making a moral *assertion*, we typically *express* certain characteristic conative psychological states or motivational stances. This is a thesis about the *pragmatics* of moral assertion, a thesis that, I will argue, is explained in central cases by the *semantics* of moral terms.⁴ It is not a thesis about the intrinsic nature of moral belief or thought; it does not imply that the state of accepting a moral judgment *consists* in part in a motivational state, or that it *entails* the existence of a motivational state. Hence, realist-expressivism is entirely compatible with *externalism*, which is the denial of internalism. Even so, as I will explain, it can do justice to many of the intuitions that fuel internalism.

Second, it is important not to confuse realist-expressivism with a sophisticated combination of antirealism and a deflationist account of the meaning of “true.” An example of deflationism is the doctrine that to call a sentence “true” is simply to affirm it. This doctrine would permit an antirealist-expressivist to affirm, consistently with his antirealism, that, for example, it is true that capital punishment is wrong. For, according to the deflationist view, to affirm this is simply to affirm that capital punishment is wrong, which obviously is something that an antirealist-expressivist can do. In addition to this deflationist account of “true,” there are also deflationist accounts of the meanings of the terms “property” and “belief.” Simon Blackburn has used the central ideas of deflationism to stake out a

3 Of course, it is committed to the first half. One subtlety that I need to ignore in this chapter is that typical versions of *antirealist-expressivism* are committed to the second half of the view. Of course, no version is committed to the first (realist) half. Typical forms of antirealist-expressivism are internalist, but realist-expressivism is not (or need not be).

4 The precise location of the line between semantics and pragmatics is controversial. The basic idea, however, is that semantics is concerned with the literal meaning of terms, expressions, sentences, and the like, insofar as their meaning can be determined independently of the contexts in which they are used. Pragmatics is concerned with properties of expressions and the like that are determined by their use, or by the contexts in which they are used. For example, the fact that the sentence “I promise to meet you” *can* be used to make a promise is a feature of its semantics. However, the question of whether a person *has* made a promise in uttering the sentence in a given context is a question in pragmatics. General questions about what a context must be like in order for a person to make a promise in uttering the sentence, and questions about what is required in order to use the sentence sincerely to make a promise, are also questions in pragmatics. I am grateful to Steven Davis for help with this distinction.

position he calls “quasi-realism,”⁵ a position that he intends to be antirealist and expressivist. Deflationism permits Blackburn to combine antirealist expressivism with the thesis that there are indeed moral “truths,” moral “properties,” and moral “beliefs.”⁶ Given that Blackburn’s is an antirealist view, my own thesis that we can combine expressivism with the idea that there are moral truths, moral properties, and moral beliefs might also be viewed as antirealist. But even though realist-expressivism does justice to the expressive characteristics of moral discourse, it is fully realist. Any version of moral realism could, in principle, be incorporated into a version of realist-expressivism. Realist-expressivism can be as realist as one would like.

Given the existence of deflationist positions of the kind I have described, however, the distinction between realism and antirealist-expressivism can seem to disappear. Realism holds that moral thoughts have truth conditions, but an antirealist-expressivist who is also a deflationist about “true” would concede that moral claims have truth conditions. It is trivial that, for example, “Capital punishment is wrong” is true just under the condition that capital punishment is wrong. The distinction between moral realism and antirealist-expressivism is subtle, but it is crucial to our understanding of the cognitive status of moral discourse. The best way to characterize the distinction, I think, is in terms of a difference between the semantics of moral predicates and the semantics of familiar garden-variety descriptive predicates, and in terms of a difference between the metaphysics of the ‘properties’ referred to by predicates of these kinds.

Put in these terms, moral realism holds that moral predicates have the same basic semantic characteristics as at least some typical nonmoral ‘descriptive’ predicates. Let us say, for convenience, that the central semantic role of the latter is to ‘refer’ to ‘properties,’ but let us do so without committing ourselves to any particular metaphysical account either of reference or of the nature of properties. Given this manner of speaking, we

5 See Blackburn 1997, 1993. Allan Gibbard exhibits a temptation toward quasi-realism as well (in Gibbard 1990). For more on Gibbard’s views, see Horwich 1993.

6 A deflationist about the term “property” might hold that to say there is a property of rightness is simply to affirm that some things are right. A deflationist about “belief” might hold that to say that a person believes that some things are right is simply to say that the person is disposed sincerely to affirm sentences to the effect that some things are right. On views of this kind, an antirealist-expressivist obviously can affirm, consistently, that there is a property of moral rightness and that there are beliefs about the rightness of actions; to affirm these claims would simply be to affirm that some things are right and that some people are disposed sincerely to affirm sentences to the effect that some actions are right.

say that the predicate “perennial” refers to the property of being a perennial. Similarly, moral realism holds, moral predicates, such as “wrong,” refer to moral properties, such as wrongness. In the first place, then, moral realism holds that the chief semantic role of the moral predicates is to refer to moral properties, such as rightness, wrongness, virtuousness, or viciousness. Second, it holds that these properties have the same basic metaphysical status as ordinary nonmoral properties, whatever that is. Simply for convenience, let me speak of these properties as metaphysically “robust.”⁷ Third, moral realism holds that ‘basic’ moral propositions are true, just as ordinary descriptive claims are true, when the relevant things have the relevant properties; it adds that some basic moral propositions are in fact true.⁸ Realist-expressivism accepts all three of these core realist claims.

Antirealist-expressivism, however, denies the first two of these claims. Deflationism would allow an antirealist-expressivist to agree that there are moral properties and that moral predicates refer to moral properties. It would also allow her to agree both that basic moral propositions are true when the relevant things have the relevant properties and that some basic moral propositions are true. However, an antirealist-expressivist would deny that the moral properties referred to by the so-called thin moral predicates, such as “wrong” and “good,” are robust. That is, she would deny that moral properties have the same metaphysical status as ordinary

- 7 As noted in the text, the problem I am addressing is how to distinguish between moral realism and antirealist-expressivism, given a deflationist account of the meaning of “true.” Hartry Field has proposed that the distinction is best drawn in terms of the idea of an “objectively correct” norm (1994, 440–441). I am proposing that the distinction can be drawn in terms of the semantic role of ordinary nonmoral predicates and the idea of a ‘robust property,’ which is in turn explained in terms of the metaphysical status of the referents of ordinary predicate terms. The issues raised by questions about this metaphysical status go beyond the scope of this chapter. The vagueness in what I am proposing is due in part to the fact that, as Michael Devitt has stressed, a formulation of the debate between moral realism and antirealist-expressivism ought to be independent of general metaphysical issues about the nature of properties (1997, 302–320, esp. 316–318). Among other things, such a formulation ought to allow for a nominalist understanding of talk of ‘properties,’ even though a nominalist would deny that there are any ‘properties’ at all under some understandings of what this would mean. This is why I speak above of “the metaphysical status of the referents of ordinary predicate terms,” and it is why, in the text, I speak of the “semantic role” of such predicate terms. I am attempting to be neutral among various accounts of these matters.
- 8 A ‘basic’ moral proposition is a proposition that entails, for some moral property M, that something instantiates M. An example is the proposition that capital punishment is wrong. Among nonbasic moral propositions are propositions such as that nothing is morally wrong and that either abortion is wrong or $2 + 2 = 4$. In Copp 1995, I called basic moral propositions “paradigmatic.”

nonmoral properties, such as the property of being a perennial. She would also deny that reference to moral properties is the chief semantic role of the moral predicates. Under antirealist-expressivism, the distinctive aspect of the meaning of moral predicates, the aspect that distinguishes *moral* predicates from other kinds of predicates, is not that they refer to a special kind of property. Their chief and distinctive semantic role, at least in their paradigmatic use in making moral judgments, is, instead, to express certain characteristic emotive or conative states of mind, such as ‘prescribing,’ ‘commending,’ or ‘expressing acceptance of a norm.’ On such an account, a person who says that something is “wrong” does not primarily assert that the action in question has the property wrongness; instead, she expresses disapproval of the action, or some other attitude toward it. Hence, according to antirealist-expressivism, the semantics of these moral predicates is quite unlike the semantics of nonmoral descriptive predicates.

To be sure, there are ‘thick’ moral predicates, such as “honest” and “kind,” and an antirealist-expressivist would concede that these predicates refer to properties that are robust in the way that ordinary nonmoral properties are robust. She would insist, however, that this is because the properties they refer to *are* ordinary nonmoral properties – psychological properties, for example – and not moral properties. Hence, she might hold that “honest” refers to a disposition to assert only what one takes to be true. Even in the case of thick moral predicates, however, she would argue that the characteristic semantic role of such predicates, at least in their paradigmatic use in making moral judgments, is to express a characteristic emotive or conative state of mind.

Antirealist-expressivists disagree about the details, of course. Nevertheless, it can be seen that antirealist-expressivism combines a negative thesis to the effect that there are no robust moral properties, and that the chief semantic role of the moral predicates is not to refer to properties, with a positive thesis to the effect that the characteristic semantic role of moral predicates is to express a distinctive conative or motivational state of mind. The negative thesis conflicts with realism, but the positive thesis is logically independent of this negative thesis. There is, therefore, room for a kind of expressivist moral realism. As a version of moral realism, this view would hold that moral predicates refer to moral properties that are metaphysically akin to ordinary nonmoral properties. As a version of expressivism, the view would hold that at least one of the semantic roles of moral predicates in their paradigmatic uses is to express a certain characteristic conative or motivational state of mind.

The availability of this sort of view is perhaps obvious, once stated. The difficulty is to develop the details in a plausible way. If we can do so, we can defuse many of the familiar arguments for antirealist forms of expressivism. It should already be obvious that an argument for the positive thesis of expressivism is not an argument for the negative antirealist thesis. As I will explain, realist-expressivism aims to do justice to many of the intuitions that fuel the familiar arguments for antirealist-expressivism. For example, realist-expressivism captures the intuition that to call capital punishment “wrong” is to express disapproval of it, a disapproval that does not consist simply in believing it is wrong. Yet realist-expressivism is entirely compatible with the realist thesis that the wrongness of capital punishment would consist simply in capital punishment’s having the robust property of wrongness. Once both of these things are understood, I think it will be clear that realist-expressivism is a genuinely interesting view. It can be accepted by moral realists even though it captures intuitions that can seem to ground antirealism.

In order fully to develop a version of realist-expressivism, one must develop both the realist side of it and the expressivist side. There are two chief issues on the expressivist side. First, what kind of conative or motivational state of mind is expressed in making a moral judgment? In principle, any answer to this question that is given by an antirealist-expressivist can also be given by a realist-expressivist. In the next section of this chapter, I briefly discuss the views of a number of prominent antirealist expressivists. The view that I propose later in the chapter, near the end of section 6, is similar to Allan Gibbard’s, according to which, as I will explain, the relevant state of mind is a state of norm acceptance. In sections 3 through 5, I focus on the second chief issue: In what sense are the relevant conative or motivational states of mind “expressed”? It turns out that answering this question leads to surprising complexities. In section 6, I propose the form of realist-expressivism that I think is most plausible, developing both the realist and the expressivist sides of the view. I specify what kind of conative state of mind is, I think, expressed in making a moral judgment. In section 7, I provide arguments for my proposal, at least for the expressivist side of it. In section 8, I briefly return to the task of exhibiting the differences between realist-expressivism and both internalist moral realism, on the one hand, and the combination of deflationism with antirealist-expressivism, on the other hand. I want to stress that my main task in this chapter is simply to propose a plausible formulation of realist-expressivism and to illustrate its theoretical advantages. The arguments for realist-expressivism are not conclusive, but I believe it is a position that deserves attention.

In order to discuss these issues, it will be useful to adopt a uniform terminology. I will use the expression “moral judgment” to refer to the kind of speech-act we perform in making a moral claim.⁹ Making a moral judgment consists in uttering a sentence with a relevant meaning with relevant intentions in a relevant context. I will use the expression “moral thought” to refer to the state of mind expressed by a person in making a moral judgment, leaving open whether such thoughts are, or involve, beliefs. For example, a person typically would be making a moral judgment in telling a man that he morally ought not to steal. But we do not need to use moral terms, such as “wrong” and “virtuous,” in order to make moral judgments. I might intend to make a moral judgment, and I might succeed in doing so, if I say, “I would have thought twice before X-ing!” In saying this, I might succeed in expressing the thought that it was wrong to X, even though the sentence I use obviously does not mean that it was wrong to X. This example shows that a distinction needs to be drawn between what a *sentence* (literally) means or implies and what a *person* means or implies in using a sentence. In speaking of the “meaning” of *sentences*, I will talk of their “content.” I will use the term “proposition” in speaking of contents that have a truth value. When the content of a sentence is (or includes) a proposition, the sentence would standardly be said to “express” the proposition, but this way of talking would invite confusion since I will be talking of assertions as “expressing” states of mind, which is a different matter. To avoid confusion, then, when the content of a sentence is (or includes) a proposition, I will say that the sentence “states” the proposition. I will also speak of imperative sentences as “stating” imperatives or commands.

Typically, when a person uses a declarative sentence with a relevant intention in a relevant context, she expresses a belief, where a belief takes a proposition as its object. For example, if I say that stealing is widespread, then in most contexts I would express the belief that stealing is widespread. In such cases, I will say that one makes “assertions.” An assertion, then, is the use of a declarative sentence to express a belief. Since I am leaving it open whether moral thoughts are or involve beliefs, I am also leaving it open whether to make a moral judgment is to “assert” something in this technical sense.

9 The classic sources of speech-act theory are Austin 1962 and Searle 1969. For helpful discussion, see Bach and Harnish 1979 and Davis 1976, 16–27.

An antirealist-expressivist can agree that some moral thoughts consist, in part, in a belief. For example, the thought that Clinton is honest includes a belief attributing the property of honesty to Clinton. According to the negative thesis of antirealist-expressivism, however, there are no robust moral properties. For the antirealist-expressivist, then, the property of honesty is not a robust *moral* property, and moral thoughts do not involve accepting propositions in which robust *moral* properties are attributed to things. As I discussed in section 1, the positive thesis of antirealist-expressivism specifies what expressivists take to be distinctive about moral thoughts; their idea is that moral thoughts consist, at least in part, in a certain characteristic kind of conative or motivational state of mind. The positive thesis also specifies that the chief semantic role of moral predicates, in their paradigmatic use, is to express such a state. The history of expressivism contains a number of proposals about the nature of this state of mind; most notable are the proposals of Charles Stevenson, A. J. Ayer, R. M. Hare, Blackburn, and Gibbard.¹⁰

Stevenson, Ayer, and Hare accept the positive thesis of expressivism as I stated it. For Stevenson and Ayer, to make a moral judgment is at least in part to express a conative attitude. For Hare, to make a moral judgment is at least in part to state a command; it is to commend something, or to prescribe or enjoin the doing of something. In my terminology, all three would agree that some state of mind other than belief is involved in having a moral thought. For Stevenson and Ayer, to have a moral ‘thought’ is at least in part to have a certain moral attitude. For Hare, it is to assent to or subscribe to a relevant universal command, which is in part to have an appropriate intention.¹¹

Contemporary expressivists have proposed similar doctrines. Blackburn proposes, for instance, that “the fundamental state of mind of one who has an ethical commitment” is best conceived as a “stance” rather than a belief because of the connection of ethical commitments with reaction, action, and choice, rather than information. To be sure, we do speak of “moral beliefs,” but, given his deflationism, Blackburn holds that this, and other features of what he calls “the important surface phenomena of ethics,” are compatible with expressivism.¹² Gibbard proposes that to call something rational is not to attribute a property to it, but rather to express

10 Stevenson 1997; Ayer 1952, 108; Hare 1952, 1–5; Blackburn 1997; Gibbard 1990.

11 Stevenson 1997, 74, 78, 79; Ayer 1952, 108; Hare 1952, 4, 13, 20, 168–172.

12 Blackburn 1997, 168–169.

a certain state of mind, the state of “thinking something rational.”¹³ He then holds that to think something rational is, roughly, to accept norms that permit it, where a “norm” is a content that is “expressible by an imperative.”¹⁴ To “accept” a norm in the relevant sense is to have a distinctive complex of attitudes and dispositions toward a norm and the actions it calls for. Norm acceptance is a kind of “motivational state.” Gibbard then explains moral states of mind in terms of norm acceptance. Hence, he says, to think an act wrong is, roughly, “to accept norms for guilt and resentment that, *prima facie*, would sanction guilt and resentment if the act were performed.”¹⁵

Realist-expressivism accepts the key positive thesis shared by these philosophers. It agrees that, for any basic moral thought that *M*, there is a conative or motivational state *C-M*, a state of some kind similar to a desire, such that a person making the judgment that *M* “expresses” the state *C-M*. It holds, however, that a person who makes a moral judgment that *M* expresses *both* the moral belief that *M* *and* a corresponding state *C-M*. Realist-expressivism combines the chief doctrines of moral realism with a central positive view of expressivism. As noted above, both the realist and expressivist sides of it need to be explained, as well as the relation between them. In the [next section](#), I begin to explain the expressivist half of the view: In what sense of “express” could it be that to make a moral judgment is to express a conative state of mind?

3. EXPRESSION, SINCERITY, AND THE PRAGMATICS OF ASSERTION

The term “express” can be used to pick out several different relations between utterances and states of mind. For example, there are causal relations of the kind that obtains between a sneer and the contempt that it expresses. Perhaps the making of a moral judgment that *M* could be caused in a similar way by the corresponding state *C-M*. However, the expressivist thesis in realist-expressivism is not simply a causal thesis. It claims that

13 Gibbard 1990, 8.

14 Ibid., 7, 46, 70. On norm acceptance, see *ibid.*, 55–57.

15 Ibid., 47. Gibbard ultimately says that normative “beliefs” are “much like any other beliefs” (100). In his fully developed view, the state of thinking an action rational is more complex than that of accepting norms that permit it. It consists, roughly, in *ruling out* all combinations of a normative system with a possible state of the world which are such that the normative system would prohibit the action in the given state of the world.

there is a *linguistically* significant relation between moral assertions and relevant conative states of mind, a relation that depends in some essential way on the fact that moral assertions are speech-acts that have moral content.

Even if we restrict attention to relations of this kind, there is more than one candidate. First, there is the relation between the assertion that *p* and the belief that *p*. The assertion “expresses” the belief, and one might propose that a person making a moral assertion that *M* expresses the corresponding state *C-M* in exactly the sense of “express” as that in which she expresses the belief that *M*. Second, there is the relation, which I think is weaker, that obtains generally between speech-acts and their sincerity conditions. This is the sense in which promises “express” intentions, and apologies “express” regret. A promise is not sincere unless the promisor intends to carry through, and an apology is not sincere unless accompanied by regret.¹⁶ One might propose that moral assertions express relevant conative states of mind in this sense of “express”; put more formally, the idea of this proposal is that the sincerity of a person making a moral judgment that *M* depends on her being in a corresponding state *C-M*. I think that neither of these proposals would be adequate by itself, for even if either were true, its truth would need to be explained. If either

16 Similarly, if I assert something, my sincerity depends on my believing what I say. “Moore’s paradox” reveals that more than just this is involved in the relation between assertion and belief. To see this, consider the Moore-paradoxical sentence, “There is a smokestack in Bowling Green but I do not believe there is a smokestack in Bowling Green.” If I utter this sentence, and if I am thereby asserting that there is a smokestack in Bowling Green, my sincerity depends on my believing that there is a smokestack in Bowling Green, which I then say that I do not believe. Hence I undermine my own sincerity. But, more than this, in uttering the Moore-paradoxical sentence, I do not succeed in asserting that there is a smokestack in Bowling Green because asserting something involves a kind of commitment to belief that I reject in the last half of the utterance. Indeed, it is not clear, other things being equal, what speech-act I perform in uttering this sentence. Compare this case with that of promising. If I promise that *p*, my sincerity depends on my intending that *p*, so it would be odd to say, “I promise to build a smokestack in Bowling Green but I have no intention of building one.” In saying this, I would undermine the sincerity of my own promise. Despite this, however, I might succeed in promising, for I might obligate myself to build a smokestack even though what I say implies that my promise is insincere. Hence, it seems, the assertion that *p* involves a commitment to believing that *p* that *cannot* be canceled without undermining the assertion. In contrast, although the promise that *p* involves a kind of commitment to intending that *p*, it appears that this commitment *can* be canceled without undermining the promise, even though canceling it does undermine the *sincerity* of the promise. See Grice 1989, 42. Bach and Harnish provide an account of assertion that elegantly explains why it is that a person who says that *p*, and then adds that he does not believe that *p*, would fail thereby to assert that *p* (1979, 15–16).

were true, we would want to find something deeper that would explain its being true.

In order to forestall misunderstanding, it is important to note right away that, understood in a certain way, it cannot be denied that a person who says, for instance, “Cursing is wrong,” expresses moral disapproval in exactly the sense in which she expresses her belief that cursing is wrong. This is because there is a straightforward sense in which simply to *believe* cursing is wrong *is* to morally disapprove of cursing. We can report a person’s belief by saying she “disapproves” of cursing. Nothing more need be true of a person who morally disapproves of cursing than that she believes cursing is wrong.¹⁷ So of course a person who says, “Cursing is wrong,” expresses her disapproval of cursing in the sense in which she expresses her belief that cursing is wrong. However, this point does not help expressivism to find a conative and motivational state distinct from belief that is expressed in moral assertion.

The thesis that is on the table, then, is that there is some conative or motivational state C-M, distinct from the belief that M, that is expressed in making the moral assertion that M. The first proposal is that this state is expressed in exactly the sense in which assertions “express” beliefs. What, though, is this sense of “express”? Fortunately, we do not need to provide a fully adequate answer to this question. This is because, if a moral assertion that M is related to a corresponding state C-M in exactly the way that the assertion is related to the belief that M, then since the assertion’s sincerity depends on whether the speaker believes that M, it follows that its sincerity must also depend on whether the speaker is in the state C-M. This, in effect, is the second proposal. That is, the first proposal entails the second, and is not adequate unless the second is adequate. We can therefore focus on the second proposal.

Gibbard invokes the distinction between the truth conditions and the sincerity conditions of an assertion to distinguish between a person’s asserting that she is in a given state of mind and her expressing that state of mind. I *assert* that I am in a state of mind in saying something just in case the *truth* of what I say depends on my being in that state of mind. In contrast, I *express* a state of mind in saying something just in case *sincerity* demands that I be in that state of mind, given what I say and assuming I know the meaning of what I say.¹⁸ Let us call expression in this sense *sincerity-expression*. Insincerity involves deceit, feigning, or pretense. For

17 Cf. Hare 1952, 10.

18 Gibbard 1990, 84.

example, there is a pretense involved if I apologize for something that I do not regret. If, in saying something, I sincerity-express a certain attitude, there would be a pretense involved in my saying the thing if I do not have that attitude.

Assertions are not unique among speech-acts in having sincerity conditions. As we saw, my sincerity in apologizing for something depends on my regretting it, and my sincerity in promising to do something depends on my intending to do it. It should be clear, however, that I can succeed in performing a speech-act of one of these kinds even if I am insincere. To say that I apologize for what I did is to apologize even if I am insincere. I can promise to do something even if I have no intention of doing it. Similarly, I can successfully assert that *p*, and thereby sincerity-express the belief that *p*, even if I do not actually believe that *p*. That is, the sincerity conditions of a speech-act are not necessarily among its success conditions.

The second proposal, then, is that the sincerity of a person making a moral judgment that *M* depends on her being in a corresponding state *C-M*. The plausibility of this idea can be illustrated with an example of Michael Smith's. Suppose you persuade me that I ought to give to famine relief, and soon thereafter I have the opportunity to make a donation. If I fail to exhibit any motivation to give, you will find my behavior puzzling. As Smith says, my behavior will "cast serious doubt on the sincerity of my claim to have been convinced that it is right to give to famine relief at all."¹⁹ This example suggests that in making the judgment that I ought to give, I sincerity-express a motivation to give. However, the example does not *show* this. What it appears to show is merely that, as Smith says, my lack of motivation casts "serious doubt" on the sincerity of my claim to believe I ought to give to famine relief. Perhaps, that is, my lack of motivation is *evidence* that I do not believe what I say, and therefore it is *evidence* that my assertion is insincere. But this falls short of showing that motivation is a sincerity condition of moral assertion. For it falls short of showing that my assertion that I ought to give when I am not motivated is *in itself* insincere in the way that it is *in itself* insincere to apologize when one has no regret. Consider an analogy. If a child blushes while saying something, this might be evidence that the child is speaking insincerely. However, to say something while blushing is not *in itself* insincere, and speaking without blushing is not a sincerity condition of assertion. Hence, I claim, even if Smith's example suggests that in making the judgment that

19 Smith 1994, 7.

I ought to give, I sincerity-express a motivation to give, it does not actually show this.²⁰

Consider a revised version of the example. Suppose that a known amoralist agrees that he ought morally to give to famine relief; he says, “I *could* refuse to give, but it would be *wrong*.” Soon thereafter he has the opportunity to make a donation but fails to exhibit any motivation to give. If we know him to be an amoralist who is quite unmoved by moral considerations, we surely will find his behavior to be in character and not at all puzzling. Suppose he said, “I agree that it would be wrong of me to refuse to give, but of course I am not at all motivated to avoid wrongdoing. I rather like wrongdoing.” We would take him to have asserted, among other things, that it would be wrong of him to refuse to give. Moreover I do not think that we would take him to have been insincere, for there was no apparent deceit or pretense on his part.²¹ Perhaps Smith is correct that there are circumstances in which a person who said, “I ought to give to famine relief,” would sincerity-express a motivation to give, but an amoralist who is open about his lack of motivation is hardly speaking under the *pretense* of being motivated to give when he says he ought to give. It might be replied that he is speaking under the pretense of *believing* he ought to give, but we can stipulate that he sincerely *believes* that he believes he ought to give. It appears, then, that there need not be any insincerity on his part. If this is correct, then the sincerity of a person making a moral judgment about what he ought to do does not depend in general on his being motivated accordingly.

Hare would object here that the amoralist in my example does not use moral terms in their standard senses. In Hare’s view, an amoralist who says, “I ought morally to give to famine relief,” but then explicitly denies that he is motivated to give to famine relief would be using the phrase “ought morally” in what Hare would call an “inverted commas” sense.²² Hence, the amoralist would not assert, nor would he believe, that he literally *ought morally* to give to famine relief – or so Hare would respond. But Hare’s response is surely implausible. It implies, for example, that the amoralist expresses a different belief in saying, “I ought morally to give to famine relief,” than you or I would express in saying this. This claim looks problematic if we consider cases in which we disagree with the amoralist. Suppose, for example, that we have a theory about the effects of

20 Janice Dowell helped me to think through Smith’s example.

21 Steven Davis urged me, in conversation, to note cases of this kind.

22 Hare 1952, 124–126, 167ff. See also Smith 1994, 68–71.

famine-relief programs that suggests it would be wrong to give to famine relief. In such a case, it would be natural for us to deny what the amoralist says. On Hare's view, however, to deny what the amoralist says would *not* be to deny that one *ought morally* to give to famine relief. This strikes me as an implausible implication of Hare's position and I therefore think that Hare's response is unsuccessful.²³

There is nevertheless something important in Hare's response. If it is viewed abstractly, Hare's claim is that the fact that a person using moral terms to make a moral judgment expresses a corresponding conative state is due to the *meaning* of moral terms when they are used literally. He is also claiming that this aspect of the meaning of moral terms can be canceled or eliminated by placing the terms in "inverted commas." I think that, so understood, Hare's contention is correct. As I will argue, however, this is compatible with moral realism. Indeed, it is compatible with externalist moral realism. To explain why I say this, I need to introduce my own account of the relation between moral judgments and conative states of mind.

For present purposes, the important thing is that both the first and second proposed accounts of this relationship are about the pragmatics of moral assertion. If either of them is true, however, the explanation surely is semantic. It lies in *what* we assert, in making moral assertions. That is, if it is true in a certain context that the sincerity of a person making the moral judgment that M depends on her being in a distinctive state of mind C-M, this must be due to something distinctive about the *content* of M. Consider the sentence "Cursing is widespread," which cannot be used literally to make a moral judgment. When this sentence is used to make an assertion, the speaker's sincerity depends on her beliefs, but not on her being in any particular conative state. According to both the first and second proposals, however, when the sentence "Cursing is morally wrong" is used to make an assertion, the sincerity of the speaker depends *both* on her beliefs *and* on her being in a relevant conative state. The explanation for this alleged difference between these two sentences would surely be due to the meaning of "morally wrong."

23 For Hare, the amoralist does not make a moral judgment. Rather, he expresses a belief about the moral judgments of other people (1952, 124), or perhaps a belief about relevant local moral standards, such as the belief that the local moral standards require giving to famine relief (167). Therefore, on Hare's view, to deny what the amoralist says would be to deny something of this kind. It would not be to make a moral judgment, and hence it would not be to judge that one morally ought not to give to famine relief. This is what strikes me as implausible. A full discussion of these matters is outside the scope of this chapter.

Pejorative terms are characterized by a feature that Gottlob Frege called “coloring.”²⁴ The coloring of a term is a characteristic of the term’s meaning, at least in a wide sense of “meaning,” for it is a characteristic of the linguistic conventions governing the use of the term. It is in virtue of the coloring of various familiar impolite terms for minority groups, for example, that their use standardly expresses contempt. The form of realist-expressivism that I want to propose holds that the meaning of moral terms also includes coloring. It is in virtue of their coloring, on this account, that moral terms standardly express relevant conative states of mind. To be more exact, on this view it is a matter of linguistic convention that in asserting a basic moral proposition by uttering a sentence in which a moral term is used, a speaker “expresses” a relevant conative state of mind, other things being equal.²⁵ According to realist-expressivism, the speaker also expresses a moral belief.

Frege says in one passage that if we compare the sentences “This dog howled the whole night” and “This cur howled the whole night,” we see that they state the same “thought.”²⁶ As he says, “a different thought does not correspond to every difference in the words used.” The second sentence “puts us . . . in mind of a dog with a somewhat unkempt appearance,” but even if the dog is actually neatly groomed, it is not the case, he says, that the former sentence is true and the latter false. They have the same truth value. “What distinguishes the second from the first,” he says, “is of the nature of an interjection.” The word “cur” implies contempt, and because of this a person might want to avoid using the word in describing a dog toward which she does not feel contempt. Despite this, Frege asks us to imagine the second sentence being “spoken by someone who does not actually feel the contempt which the word ‘cur’ seems to imply.” Even in this case, Frege says, if the first sentence is true, “the use of the word ‘cur’ does not prevent us from holding that the second sentence is true as well.” For, he says, “we have to make a distinction between the thoughts that are expressed and those which the speaker leads others to take as true although he does not express them.” A speaker who called the howling dog a “cur” would imply that she feels contempt for the dog, but

24 Frege 1984c, 161; 1984b, 185; 1984d, 357. See also Frege 1979c, 140–141; and 1979a, 197–198. I owe these references to Janice Dowell and Kent Bach.

25 Recall that a basic moral proposition is a proposition that entails that something instantiates M, where M is a moral property. See note 8 above.

26 Frege 1979c, 140–141. This paragraph and the next follow Frege’s discussion.

this is analogous, says Frege, to the fact that “a speaker who announces the news of a death in a sad tone of voice” creates the impression that he is sad. In such a case the speaker does not state the proposition that he is sad, or say he is sad, even though he displays sadness. “Naturally things are different,” Frege writes, when a difference of the kind we are discussing is a matter of “common usage,” as is presumably the case with the difference between “dog” and “cur.”

Frege would deny that there is any difference in the reference or the sense of the terms “dog” and “cur.” This is perhaps not true to the meaning of “cur” in English – it is defined as “mongrel dog” in at least one dictionary²⁷ – but we should not worry too much about the example. Let us assume that there is no difference between the references or senses of the terms “mongrel dog” and “cur.” That is, let us assume that there is no difference in the properties referred to by the terms or in the concepts they express. Given this assumption, Frege would say there is no difference in the thoughts or propositions stated by the sentences “This mongrel dog howled all night” and “This cur howled all night.” Yet, Frege would insist, there is a difference in the “contents” of the sentences. He holds that the proposition or thought stated by a sentence is only a “part” of its content, “the part that alone can be accepted as true or rejected as false.” He says, “I call anything else that goes to make up the content of a sentence the colouring of the thought.”²⁸ Because of differences in coloring, there is a difference between the speech-act performed in saying that a “cur” howled all night and the speech-act performed in saying that a “mongrel dog” howled all night. There is a difference in what speakers of each sentence would imply about their states of mind because there is a difference in the linguistic conventions governing the uses of the terms. Following Frege, I will speak in such cases of differences in coloring. “Cur” has a coloring such that its use standardly conveys contempt; “mongrel dog” does not have such a coloring.

The difference in coloring between “cur” and “mongrel dog” is a matter of “common usage.” I want to say it is a difference in the “meaning” of the terms because the different conventions for their use affect what is *communicated* or *conveyed* by uses of the terms. I suppose that one might want

27 *American Heritage Dictionary of the English Language*, 3rd ed., s.v. “cur.” Kent Bach and Thomas E. Hill, Jr., drew my attention, in conversation, to worries about whether the meaning of “cur” in contemporary English undermines the effectiveness of Frege’s example.

28 Frege 1979a, 197–198. I think it would be preferable to say that the coloring is a property of the sentence used to state the thought rather than a property of the thought itself.

to restrict the use of “meaning” for certain theoretical purposes such that coloring is not treated as a matter of “meaning.” For my purposes, it will be enough to acknowledge that the coloring of a term can be due to the linguistic conventions governing its use. For convenience, I will say that the expressions “cur” and “mongrel dog” have the same “core meaning” even though they have different “meanings” in a wide sense of the term.²⁹

We now have an example of an assertion that is plausibly taken to express both a belief and a conative state. Assertions of “This cur howled all night” would express both the relevant belief and an attitude of contempt. There are numerous other familiar contemporary examples of colored terms expressing contempt. “Yankee” is used in a variety of countries to express contempt for Americans. “Canuck” is used in New England to express contempt for Canadians, particularly Francophone Canadians. Another example is the term “redneck.”³⁰

Inasmuch as the core meaning and coloring of a term are a function of the conventions governing its use, and since conventions can be of different kinds and can change over time, a variety of features of a term’s meaning in the wide sense can be aspects of its coloring, and the distinction between core meaning and coloring can be fuzzy. As Frege says

Of course borderline cases can arise because language changes. Something that was not originally employed as a means of expressing a thought may eventually come to do this because it has constantly been used in cases of the same kind. A thought which to begin with was only suggested by an expression may come to be explicitly asserted by it. And in the period in between different interpretations will be possible. But the distinction itself is not obliterated by such fluctuations in language. In the present context the only essential thing is that a different thought does not correspond to every difference in the words used, and that we have a means of deciding what is and what is not part of the thought, even though, with language constantly developing, it may at times be difficult to apply.³¹

What means do we have for deciding which aspects of a term’s meaning in a wide sense are aspects of its coloring?

Suppose that a person says, “Your cur howled all night.” As I said, Frege holds that even if the person “does not actually feel the contempt

29 Michael Dummett would say, I think, that the relevant difference between “cur” and “mongrel dog” is to be accounted for in “the theory of force,” which he takes to be part of the theory of meaning along with the theory of reference and the theory of sense. See Dummett 1993, 40, 87.

30 Cf. Stevenson on the expressions “elderly spinster” and “old maid” (1997, 77).

31 Frege 1979c, 141.

which the word ‘cur’ seems to imply,” what she says might well be true, for the dog might well have howled all night.³² This example illustrates one test we can use to determine which features of a terms’ meaning are aspects of its coloring. Call it the *truth test*: If the use of a term T in a given sentence suggests or implies that *p*, and if this is due to the coloring of T, then the belief expressed by a person asserting the sentence might be true even if *p* is false. For instance, the belief expressed by a person in saying, “Your cur howled all night,” might be true even if the person actually feels no contempt for your dog.

Despite this, however, it would be a *misuse* of a colored term to use it when the implication carried by its coloring is known or believed by the speaker to be false. This is because the coloring of a term is a feature of its meaning in a broad sense; it is a feature of the linguistic conventions governing the term’s use. Accordingly, it would be *inappropriate* to use a colored term in a context in which such use conveys that *p* when the speaker knows or believes that *p* is not the case, even if doing so does not result in what the speaker says being false. Not all inappropriateness of use involves speaking falsely. A speaker who says, “Your cur howled all night,” might speak truly even if she lacks contempt for the dog in question, but she does nevertheless speak inappropriately. For one thing, her usage is potentially misleading. But even if she attempts to cancel the suggestion of contempt by saying, “Your cur howled all night but I do not mean to imply that I have contempt for your dog,” she might find herself being corrected or challenged. Her use of the term “cur” calls out for explanation. It is linguistically inappropriate.

Still, it seems to me that the implications carried by coloring are *cancelable*. To be sure, they are not cancelable in the precise sense that was intended by Paul Grice, who introduced the notion that certain kinds of implication are cancelable. Grice said the implication that *p* is “explicitly cancelable” if it would be “admissible to add *but not p*, or *but I do not mean to imply that p*.” It is “contextually cancelable if one can find situations in which the utterance of the form of words would simply not carry the implicature.”³³ In the cur example, the person could say, “Your cur howled all night but I do not mean to imply that I have

32 Ibid., 140.

33 Grice 1989, 44, 39 (emphasis in original). Grice holds that “conversational implicatures” are cancelable but “conventional implicatures” are not. I am not disagreeing with Grice, even though I think coloring is an example of conventional implicature, for I am using the term “admissible” in a less strict sense than Grice does. (This note was revised in 2006.)

contempt for your dog.” Let us call this sentence “H.” It would be an odd sentence to utter, for doing so would involve a misuse of “cur” that might be challenged. In this sense it would *not* be “admissible” to utter H, so Grice presumably would hold that the implication of contempt is *not* cancelable in his sense. In another sense, however, it *would* be “admissible” to utter H. It would not be self-contradictory, and it would be fully intelligible as an assertion despite the nonstandard use of the term “cur.” A person who uttered H would cancel the suggestion of contempt in the sense that the hearer would not be justified to infer on the basis of what the speaker said that the speaker has contempt for the dog in question. One might think that the use of “cur” in such a context could not be literal, but this seems incorrect, because we are imagining the term to be used with at least its standard core meaning. We might wonder why a person who knows the meaning of the word “cur” would use it to speak of a mongrel dog toward which she has no contempt. The answer might be that she does not know or cannot think of the word “mongrel.” This supposition helps us to imagine a context in which a speaker says, “Your cur howled all night,” without there being any implication of contempt. Consider a situation in which you know both that the speaker likes your dog and that she does not know the word “mongrel” because she is just learning the language. In this case, the speaker could be quite sincere in saying what she says even though she has no contempt for your dog. The implication of contempt would be contextually canceled.

Cancelability is related to a more general feature we could call *alterability*. The usual implication of a colored term can be canceled or altered if what a speaker asserts is more complex than a simple sentence. It can be canceled or altered, for example, in cases where simple sentences involving colored terms are embedded in larger constructions. As we have seen, the usual suggestion of using the term “cur” is canceled if a person says, “Your cur howled all night but I do not mean to imply that I have contempt for your dog.” In cases of this kind, the second conjunct cancels the implication, otherwise carried by the speaker’s use of the term “cur,” that she feels contempt for your dog. Similarly, the remark “If your dog is a cur, you ought to sell it” does not imply that the speaker has contempt for your dog, although it does seem to imply contempt for mongrels. The explanation for this lies partly in the content of the conventions governing the uses of terms in virtue of which they have coloring. For instance, the convention governing “Yankee” could be such that *any* use of the term

expresses contempt for Americans, or it could simply be such that *calling* someone a Yankee expresses contempt for that person. Suppose someone says, "If Alice is a Yankee, she will celebrate the fourth of July." On neither convention would the person thereby express contempt for Alice, but on the former convention, the person would thereby express contempt for Americans.³⁴

Colored terms have an additional feature that was discussed by Grice, namely, that the suggestions carried by coloring should be *detachable*.³⁵ The implication of contempt carried by "cur" is detachable in the sense that the beliefs expressed by sentences containing the word can also be expressed by sentences that differ only by containing the phrase "mongrel dog" in place of the word "cur." In general, if a speaker's assertion of a sentence S implies that *p*, and if this implication is due to the coloring of a term T that is contained in S, then it would be possible in principle for the speaker to say or assert the same thing as is asserted by S, or to express the same belief as is expressed by means of S, without implying that *p*. In the cur example, the person could say, "Your mongrel dog howled all night," instead of "Your cur howled all night." Of course, in any given case, it might turn out that there is no other term or phrase in the language that has exactly the same core meaning as the colored term T. Even in that case, however, the relevant implication would still be detachable in principle.

We now have four rough tests for coloring: Frege's truth test, the misuse test, my test of cancelability, and Grice's test of detachability. If a speaker's assertion of a sentence S implies that *p*, and if this implication is due to the coloring of a term T contained in S, then (1) the belief expressed by the speaker in asserting S might be true even if *p* is false, and the

34 Kent Bach suggested to me in conversation that it might be useful to distinguish between two kinds of pejorative terms. There are (a) terms, such as "Yankee" and "cur," that are used to refer contemptuously to a class of persons or things such that their use typically expresses or implies contempt for all persons or things in that class, and (b) terms, such as "jerk," that are used to refer contemptuously to persons or things such that their use implies that the speaker has contempt for the person or thing explicitly referred to, but does not imply that she has contempt for anyone or anything else. The remark, "Alice is a Yankee," implies that the speaker has contempt for Americans in general as well as for Alice, but the remark, "Alice is a jerk," only implies contempt for Alice. Compare "If Alice shows up at the Fourth of July celebration, she is a jerk" with "If Alice shows up at the Fourth of July celebration, she is a Yankee."

35 Grice uses the notion of detachability to distinguish between "conventional" and "non-conventional" implicatures. Grice 1989, 39, 43–44.

speaker's implication of the proposition that p should be (2) detachable and (3) cancelable. Furthermore, (4) it would be a misuse of the term T for a speaker to assert S when she knows or believes that p is not the case.

We can apply these tests to the term "Yankee." Let us suppose that, in Bob's dialect, if Bob says, "Alice is a Yankee," he thereby would standardly express contempt for Alice. Despite this, (1) the belief he expresses might be true even if he has no contempt for her, for his belief would be true if Alice were an American. (2) The implication of contempt is detachable in Grice's sense, for Bob could express his belief about Alice by saying, "Alice is an American." In this way, he could avoid expressing contempt for her. Moreover, (3) he could cancel the implication that he feels contempt by saying, "Alice is a Yankee but I feel no contempt for her." Perhaps Bob lives in a community in which there is such contempt for Americans that there is no term in common use for referring to Americans except the pejorative term "Yankee." If this were so, and if it were known to be so, and if it were known that Bob is not in fact contemptuous of Americans, he would not express contempt in calling Alice a "Yankee." The implication of contempt would be contextually canceled. Despite this, (4) the conventions governing the use of "Yankee" are such that, we imagined, when used correctly and literally, "Yankee" is used to refer contemptuously to Americans. Hence, if Bob has no contempt for Alice, it would be a misuse of the term for him to call her a "Yankee." Yet Bob could be quite sincere in saying, "Alice is a Yankee," even if he feels no contempt for her. This might be so even if he knows the coloring of the term "Yankee," for perhaps he does not know any nonpejorative term that he could use in place of it.

Let us say that a speaker's use of a term *Frege-expresses* a state of mind just in case it is a matter of the term's coloring that, other things being equal, its use conveys that the speaker is in that state of mind. To be more exact, a speaker's use of a term *Frege-expresses* a state of mind just in case it is a matter of linguistic convention governing the use of the term that, other things being equal, if a speaker asserts a simple isolated subject-predicate sentence in which the term is used literally,³⁶ the fact that the speaker used the term conveys that the speaker is in the state of mind. We could call such conventions *expressive conventions*. An example might be

36 I shall not attempt to specify exactly which sentences these are.

the convention whereby calling someone a “Yankee” expresses contempt, other things being equal. An expressive convention regarding the use of a term is a convention such that speakers who use the term in contexts of a certain kind express a given conative or emotive state of mind, other things being equal. Frege-expression depends on the existence of such conventions.

The form of realist-expressivism that I want to explore holds that moral terms have coloring in virtue of which, other things being equal, their use in making typical moral assertions Frege-expresses a characteristic conative state of mind. To be more exact, the view is that it is a matter of linguistic convention governing the use of moral terms that, if a speaker asserts a basic moral proposition by using a sentence in which a moral term is used literally, the speaker implies, other things being equal, that she is in a relevant conative state of mind. Moral terms are governed by expressive conventions.

It would be premature to apply our four tests for coloring to moral terms without saying something about the nature of the kind of conative state of mind that I think is expressed by a speaker in making a moral judgment. I will claim that the state of mind in question is that of subscribing to a relevant norm, but I need to explain what I mean by this. The view I will propose is that moral terms have coloring in virtue of which, roughly, their use standardly expresses the speaker’s subscription to a relevant norm.

5. COLORING AND CONTENT

It might seem ad hoc and unhelpful to postulate the existence of otherwise unexpected linguistic conventions to explain the coloring of pejorative terms and the means by which they are used to Frege-express states of mind. It might seem worse than unhelpful to view moral terms as a species of pejorative term, which is perhaps how my proposal will be viewed at this point. For Frege, however, coloring is a characteristic of a great deal of our discourse, not just of pejorative discourse. Indeed, Stephen Neale has proposed that even proper names have coloring.³⁷ In this section of the chapter, I shall place the ideas of coloring and Frege-expression in a general theoretical framework and discuss certain objections that derive from issues in philosophy of language.

37 Neale 1999, 72–73. I owe this reference to Kent Bach.

Neale thinks that coloring “may constitute only the tip of a semantic iceberg.” The iceberg in question is a phenomenon that is, he says,

quite natural once we take into account the nature of communication. We do not seek to transmit information only about the world; communication may also involve the transmission of information about our attitudes and emotions; thus we convey information using expressions such as ‘It is raining’ and also sentences such as ‘Damn, it’s raining’, ‘I think it’s raining’, and ‘Damn, I think it’s raining’. That is, in many cases we use simple sentences to express a single proposition and we use modifications of those sentences to express the original proposition . . . together with a second (third, . . .) proposition.³⁸

Neale’s idea is that the uttering of a sentence may “express” or communicate a sequence of propositions, including a “primary” proposition, as well as one or more “secondary” propositions, where not all of the propositions communicated in this way are taken in every case to bear on the truth of what is thereby said.³⁹ For example, a speaker who says, “Damn, it is raining,” might communicate both the primary proposition that it is raining and the secondary proposition that he is angry or frustrated that it is raining. The truth value of what he thereby says would depend only on whether it is raining, but it would be inappropriate in most contexts for him to communicate that it is raining by saying, “Damn, it is raining,” unless he were angry or frustrated. Neale has proposed that this *multiple propositions framework* can be used to explicate a variety of phenomena, including coloring.

Consider the sentence “Smith’s cur howled all night.” Given the meaning of “cur,” a speaker uttering this sentence in a typical context presumably would communicate something like the following sequence of propositions: [primary: that Smith’s mongrel dog howled all night]; [secondary: that the speaker has contempt for mongrels, or for Smith’s mongrel]. The truth value of what the speaker says would depend only on whether Smith’s mongrel actually howled all night, but it would be semantically inappropriate for the speaker to use the sentence to state this proposition unless he had the contempt that is expressed.

38 Ibid., 60–61.

39 Ibid., 75 and throughout. See Bach 1999. Bach also introduces a multiple-propositions framework. This use of the terms “primary” and “secondary” is his. The multiple-propositions framework is quite flexible. Neale holds that the context in which a sentence is uttered, and the issues that are central in the conversation, can affect whether the falsity of a secondary proposition would lead us to view a speaker’s assertion as false. Neale 1999, 75.

In this framework, the phenomenon that Paul Grice called “conventional implicature” would be treated in the same way as the phenomenon of coloring.⁴⁰ For example, according to Grice, a person who utters, “Alice is poor, but she is honest,” using the sentence literally to make an assertion, would assert that Alice is poor and honest. Given the meaning of “but,” however, Grice thinks that the speaker would also implicate that there is some sort of contrast between poverty and honesty. The proposition that there is such a contrast is implicated rather than entailed, for even if there were no relevant contrast, what the person said would still be true if Alice were both poor and honest. Grice calls such implicatures “conventional” because, in Neale’s words, they are due to “the linguistic conventions governing the uses of the words in question.”⁴¹ In Neale’s framework, the person uttering “Alice is poor but she is honest” would communicate the following sequence of propositions: [primary: that Alice is poor and she is honest]; [secondary: that there is some relevant contrast between poverty and honesty]. The truth value of what the speaker thereby says would depend only on the truth of the primary proposition, but it would be inappropriate for the speaker to use the sentence to state this primary proposition unless she believed that the secondary proposition was also true.

Because coloring and conventional implication are treated the same way in the multiple propositions framework, I shall say that the secondary proposition communicated by the assertion of a sentence involving coloring is “conventionally implicated.” Hence, I shall say, a speaker who says, “Smith’s cur howled all night,” conventionally implicates that he has contempt for Smith’s dog. For my purposes, however, nothing turns on whether coloring is an example of conventional implicature or simply a phenomenon that is similar to conventional implicature.

In Neale’s framework, then, realist expressivism includes the thesis that, other things being equal, a person stating a basic moral claim that M using moral terminology thereby communicates the following sequence of propositions: [primary: that M]; [secondary: that the speaker is in the corresponding state C-M]. The truth value of what the speaker thereby says would depend only on whether M is the case, but it would be inappropriate for the speaker to use the sentence to state that M unless she were in state C-M.

40 On this point, see Grice 1989, 41, 46, 86; Bach 1999; and Neale 1999, 53–61.

41 Neale 1999, 53.

The multiple propositions framework allows us to see quite clearly where there might be controversy. There are at least three significantly different views concerning the alleged communication of a “secondary proposition” about the speaker’s motivations or conative state in the moral case. (1) On the first view, no propositions to the effect that the speaker is in state C-M are typically communicated or implied by a person in making a moral judgment that M. (2) On the second view, a speaker making a moral assertion that M does typically communicate a proposition to the effect that she is in a state C-M, and this proposition is among the truth conditions of what she says. According to this position, a person who judges that something is wrong asserts (perhaps among other things) that she is in a corresponding state C-M. The proposition that she is in such a state is entailed by what she says or asserts, so what she says is false if she is not in such a state. This view is therefore a kind of subjectivist relativism. (3) On the third view, the realist-expressivist view, a speaker asserting a moral proposition that M typically communicates a proposition to the effect that she is in a state C-M, but this proposition is not among the truth conditions of what the speaker says. On a view of this kind, the secondary proposition that the speaker is in state C-M is “implicated.” It is communicated, but it is not entailed by what the speaker asserts or says in making the judgment that M.

There are at least three views of this third kind, so there are three kinds of realist-expressivism. (3a) One might hold that the speaker *conversationally implicates* that she is in state C-M. When something is conversationally implicated, the implication can be canceled without any hint of semantic or linguistic oddity and without any misuse of the terms used by the speaker. Conversational implicatures are due to pragmatic features of assertions, such as assumptions standardly made by hearers.⁴² In Smith’s example, for instance, I agree that I ought to give to famine relief, and it might seem that I thereby implicate that I am motivated to give. If I do implicate this, and if the implicature is conversational, it relies on the hearers’ assumptions about our conversation. For example, it might depend on the assumptions that we were talking about what to do and that, in saying I ought to give, I was expressing a decision. In contrast to this view, (3b) one might hold that the implication that the speaker is in state C-M is “conventional” in cases in which the speaker uses moral terms in asserting that M. That is, in such cases, the proposition that the speaker is in state

42 On conversational implicature, see Grice 1989, 22–57. See also Bach and Harnish 1979, 62–64; Bach 1999, 327; and Neale 1999, 53.

C-M is communicated by the speaker in virtue of expressive conventions governing the use of moral terms, but it is not entailed by what she says in making the judgment that M. I am proposing a view of this kind. More specifically, as discussed in the [previous section](#), I am proposing that moral terms have a coloring such that a person making a moral judgment that M using moral terms Frege-expresses a corresponding state C-M.

Finally, (3c), one might agree that a speaker making a moral judgment that M typically communicates that she is in a state C-M, but hold that this is a case neither of conversational implicature nor of conventional implicature. This position is analogous to Grice's position regarding assertion and belief. Grice holds that a speaker who asserts that *p* expresses her belief that *p*, and she conveys that she believes that *p*, but he denies that the latter proposition is something she conversationally implicates.⁴³ I will say, for convenience, that it *is* something she "implicates." But the implicature is neither conversational nor conventional. It is not conversational because it cannot be canceled without oddity, and it is not conventional because it is not detachable.⁴⁴ Similarly, one might hold, a person making a moral judgment that M implicates that she is in a corresponding state C-M, but the implication is neither conversational nor conventional.

In summary, then, the alternatives are, respectively, (1) the *no C-M proposition view*, (2) the *entailment view*, (3a) the *conversational-implicature view*, (3b) the *conventional-implicature view*, and (3c) the *neither-of-the-above view*.⁴⁵ View (2), as noted above, is a version of subjectivism. It is undermined by the example of the amoralist who says he ought to give to famine relief, for we do not take the fact that the amoralist is not in a relevant conative or motivational state to *falsify* what he says about giving to famine relief. That is, we do not take his lack of motivation to show that what he says, in saying he ought to give, is *false*. I shall therefore set aside the entailment view. In section 7 of this chapter, I will argue against view (1). I will argue there that a speaker does implicate a C-M proposition in making a moral judgment. For instance, I shall argue that the person in

43 Grice 1989, 42.

44 Moore's paradox shows that we cannot successfully assert that *p* while canceling the implication that we believe that *p*. And we cannot detach the implication that we believe that *p* by carefully choosing the words we use to assert that *p*. See Grice 1989, 42.

45 It can be difficult to categorize a philosopher's view of these matters. Some of the things said by Jamie Dreier suggest, for example, that his version of "speaker relativism" is an entailment view (1990). However, the better interpretation is surely that it is either a conversational-implicature view or conventional-implicature view. It is also possible that he holds a view of kind (3c).

Smith's example who asserts that he ought to give to famine relief *implies* that he is motivated to give; he does not merely give evidence that he is motivated. The example of the amoralist suggests, however, that it is possible to cancel such implications. This is because the amoralist would apparently cancel any implication that he is motivated to give to famine relief if he said, "I agree that it would be wrong of me not to give to famine relief, but do not be misled. I am not at all motivated to give." The example therefore seems to count against the neither-of-the-above view (3c), which does not make room for the possibility of canceling the implication. For these reasons I think that the interesting controversy is between views of kinds (3a) and (3b).

In earlier work, I proposed a view of kind (3a), according to which C-M propositions are *conversationally* implicated.⁴⁶ As I will argue in section 7, however, I now think that when a speaker uses moral terms in making a basic moral assertion, the implied C-M propositions are *conventionally* implicated. In such cases, I think that the implicature reflects the *coloring* of the moral terms.

My proposal is controversial, however, in light of two arguments found in the recent literature on conventional implicature and expressivism. One is a general argument by Kent Bach that the idea of conventional implicature is a myth.⁴⁷ The second is an argument by Frank Jackson and Philip Pettit that expressivism collapses into a crude and unacceptable form of subjectivism, which in the present context amounts to an argument that the conventional implication view collapses into the entailment view.⁴⁸ If either of these arguments is sound, then realist-expressivism is only viable in the familiar guise of the conversational-implicature view. Unfortunately, a thorough discussion of the issues raised by these arguments is beyond the scope of this chapter.

The root of the arguments, however, is a challenge to philosophers who think there is a distinction between conventional implicature and entailment. The challenge is to explain how it could be that a person uttering a sentence *communicates* a proposition, in virtue of linguistic conventions governing the literal use of the terms in the sentence, without thereby *asserting* that proposition such that the truth of what the person asserts or says depends on the truth of the proposition. Suppose it is a semantic error of some kind in Bob's linguistic community to call Alice

46 Copp 1995, 35.

47 Bach 1999. However, Bach does not consider coloring. See *ibid.*, 332 n. 8.

48 Jackson and Pettit 1998. I owe this reference to Kent Bach.

a “Yankee” if one lacks contempt for Americans. How could this be so if the having of contempt is not *entailed* by what speakers say in calling people “Yankees”?

I think that the outline of the answer is clear, but I will not attempt to fill in the details. In outline, the answer is that there are different kinds of linguistic conventions. Some conventions determine what a term is to be used to refer to, when used literally. Other conventions determine that certain conative or motivational states of mind are communicated or conveyed by uses of the term in certain kinds of context. In Bob’s dialect, the predicate “Yankee” refers to Americans, but, by convention, it is also used to express contempt for Americans. Because of this, it is a misuse to use it to refer to Americans unless one has contempt. In virtue of the convention about reference, it would be false to say of Alice that she is a Yankee unless she is an American, but in virtue of the expressive convention, it would be a linguistic gaffe, “semantic offense,” or violation of “semantic proprieties” of some kind not “touching truth value” to say of Alice that she is a Yankee if one does not have contempt for Americans.⁴⁹ I think that there are both kinds of linguistic convention.

We are familiar with pragmatic conventions governing speech-acts of various kinds. According to the conventions governing promising and asserting, it is appropriate for us to promise or to assert only when (we believe) the relevant sincerity condition is fulfilled. It is clearly not the case, however, that a person making a promise reports that he intends to follow through, or that a person making an assertion reports that he has the corresponding belief. A person is taken to have promised even if she lacks the relevant intention, and a person’s assertion might be true even if she does not believe that it is. Hence, there are conventions concerning the sincerity of speech-acts that do not concern the truth conditions of what is said. The only issue is whether conventions not concerning truth conditions can be encoded in the meaning of terms, even if only in a wide sense of meaning that includes coloring. It seems to me that they can be.

6. A PROPOSED REALIST-EXPRESSIVISM

In the previous two sections, I have been explaining the sense in which, on my proposal, a person who uses moral terms to make a basic moral assertion that M standardly “expresses” the corresponding state C-M. I

49 The quoted phrases are from Grice 1989, 362, 365.

propose that she “Frege-expresses” the state C-M; that is, other things being equal, she conventionally implicates that she is in the state C-M in virtue of the coloring of her terms. Even if a person asserts a basic moral proposition without using moral terms, I still want to say that, other things being equal, she conversationally implicates that she is in the corresponding state C-M.

In order to finish introducing my proposal, I need to explain what, in my view, the emotive or conative state is that a person expresses in making a moral judgment that M. I also need to say something about the truth conditions of moral propositions in order to make clear that realist-expressivism is a form of realism as well as a form of expressivism. In this section, I will fill the most important remaining gaps on both the expressivist and realist sides of my proposal. My account will be sketchy, but I have provided more detail elsewhere.⁵⁰

Begin with the propositions stated by moral claims. Consider the claim that cursing is wrong. It would be contradictory, I submit, to say, “Cursing is morally wrong but no relevantly justified moral norm or standard prohibits it.” This would be contradictory because the wrongness of cursing entails that some justified moral standard prohibits it. By a “standard” or a “norm,” I mean a content that can be stated by an imperative; my usage here follows Gibbard’s. To a first approximation, then, my idea is that the claim that cursing is wrong states a proposition that is true only if some relevantly justified or authoritative moral standard or norm prohibits cursing. The idea is that the truth conditions of basic moral propositions are given by propositions about what is called for by relevantly justified or authoritative moral standards.⁵¹ This idea is meant to capture the normativity of moral judgment. I have called the generalization of this idea the *standard-based account* of normative judgment.

The account leaves it open what would make a moral standard relevantly authoritative. Different metaethical theories in effect make different proposals about this, and the standard-based account is amenable to various

50 I have elaborated and defended the proposal I present in the next few paragraphs in Copp 1995. For a brief introduction to this position, see Copp 1997. For a reply to some objections, see Copp 1998. See also Copp 1996 and the introduction to this book.

51 This position raises questions about the individuation of propositions and beliefs, and about the nature of philosophical analysis, that are beyond the scope of this chapter. In the text, I try to finesse these issues. For discussion, see King 1998. Recall that a basic moral proposition is a proposition that entails that something instantiates M, where M is a moral property. Something that is a standard or norm in my sense need not be embedded in the culture, nor need it be anything that people actually pay attention to in deciding how to live. A standard is the practical analogue of a proposition.

possibilities. For instance, a nonnaturalist might think that some standards simply have a “fittingness” to the moral nature of things. My own view on this underwrites a kind of naturalistic moral realism. A moral standard is relevantly justified, I suggest, just in case, roughly, its currency in the social code of the relevant society would best contribute to the society’s ability to meet its needs – including its needs for physical continuity, internal harmony and cooperative interaction, and peaceful and cooperative relations with its neighbors. I call this view the *society-centered theory*. The moral standards with currency in a society form the social moral code of the society, a system of moral standards or rules. Let me use the phrase “S-ideal moral code” to speak of the moral code the currency of which in a society S actually would best contribute to the society’s ability to meet its needs. On the society-centered theory, moral claims are true or false depending on the content of the ideal moral code for the relevantly local society. The proposal treats moral properties as relational: If cursing is wrong, it is wrong in relation to a society the ideal moral code of which would prohibit cursing. Wrongness – wrongness in relation to society S – is, roughly, the property of being prohibited by the S-ideal moral code.

The society-centered account raises issues that are not raised by the standard-based account considered by itself. For example, which society is the relevant one in any given case? What, in detail, are the needs of societies? Do different societies have significantly different needs? I will have to put such questions to one side. In previous writings, I have introduced clarifications, qualifications, and amendments to the basic idea of the theory in order to deal with these and a variety of other questions and objections.⁵²

Assuming that we have identified the relevantly authoritative moral standards, the standard-based account leaves it open precisely which proposition about these standards gives the truth conditions of a given moral proposition. As I said, the idea is that, to a first approximation, a basic moral proposition is true only if a corresponding moral standard or norm is relevantly justified or authoritative. For example, cursing is wrong only if (roughly) a moral standard prohibiting cursing is relevantly justified. In order to provide a complete account of the truth conditions for moral propositions, a great deal of detail would be required. We would need to distinguish, for instance, between the property of being morally wrong and that of being unthinkable, and between the property of being

52 See the references cited in note 50 above.

virtuous and that of being admirable. These details are not important for present purposes.

The idea that *is* important for present purposes is that a person making a moral claim expresses a moral belief, the truth conditions of which could be specified in the terms of the standard-based account and the society-centered theory. It is not a part of my view that a linguistically competent person using a given sentence to make a moral claim would know the truth conditions of the belief she expresses as they would be specified by my view. A linguistically competent person could, for instance, deny the society-centered theory without making any logical mistake. Yet I do think that a competent speaker would understand that the standard-based account is correct as far as it goes. She would realize that it would be contradictory to say, for example, “Cursing is morally wrong but no relevantly justified moral norm or standard prohibits it.” In any event, the realist and naturalist side of the realist-expressivist view I am sketching in this chapter is captured by the combination of the standard-based account and the society-centered theory.

The expressivist side of the view is the idea that a person making a moral judgment typically expresses a distinctive conative or motivational state of mind as well as a moral belief. What is this state of mind? This, recall, is a question I posed near the beginning of the chapter and promised I would return to.

Given my proposal regarding the truth conditions of the beliefs expressed by persons making moral claims, it is natural to think that a person making a moral claim also expresses her acceptance of a corresponding standard. For example, if a person believes that cursing is morally wrong, her belief is true just in case cursing is a violation of an authoritative standard that prohibits cursing. In this case my view is that, in expressing her belief, she also expresses her acceptance of a standard that prohibits cursing. Here we can make use of ideas that are found in Gibbard as well as in Hare. As noted in section 2, Gibbard writes about a state of mind he calls “norm acceptance,” which he says is a distinctive kind of motivational state. Hare writes about “accepting an imperative,” which he says is a matter of having relevant intentions. I myself have written about “subscription to a moral standard,” which I have attempted to explain as involving a syndrome of attitudes, including an intention or policy of complying with the standard.⁵³ Without attempting to define subscription

53 See Copp 1995, 84.

to a standard, let me consider the idea that a person who makes a moral judgment, and thereby expresses belief in a basic moral proposition, also expresses subscription to a corresponding moral standard. I have used the term “moral conviction” to refer to the complex state of mind that combines moral belief with subscription to a corresponding standard. The idea, then, is that a person making a moral judgment typically expresses a moral conviction.

One attractive feature of this idea is that, in Gibbard’s words, it makes explicit a “kind of endorsement” that is involved in normative discourse. This characteristic kind of endorsement is not a matter simply of having a vanilla pro-attitude or con-attitude. It is a rather more complex state of mind consisting in taking something to fall within the purview of a norm that one endorses. Gibbard therefore proposes that I have endorsed an action in the relevant sense if I have said something that expresses my acceptance of a norm that requires or permits it.⁵⁴

Another attractive feature of the idea is that it connects the state of mind of a person making a moral judgment with states that enter into her decision-making. Michael Bratman has proposed that intentions are states that constrain and guide our planning.⁵⁵ It is natural to think that a person’s moral convictions likewise constrain and guide her planning. On my view, a person who makes a moral judgment expresses her subscription to a relevant standard, which is to say, among other things, that she expresses her intention or policy to conform to such a standard. On Bratman’s planning theory, this is to say that a person making a moral judgment expresses a state of mind that constrains and guides her planning. Hence, if a person judges that cursing is morally wrong, she expresses a belief that is true only if cursing violates a relevantly warranted standard that prohibits cursing. On my view, she also expresses subscription to such a standard. If she actually does subscribe, she is in a state of mind that, if effective, constrains and guides her planning so that she is motivated to some degree not to curse.

7. ARGUMENTS FOR THE EXPRESSIVISM IN REALIST-EXPRESSIVISM

In this section, I shall investigate arguments for the expressivist half of my proposed version of realist-expressivism. The distinctive part of the

54 Gibbard 1990, 33.

55 Bratman 1987.

expressivist half of the view is that moral terms have coloring in virtue of which, other things being equal, their use in making typical moral assertions Frege-expresses a characteristic conative state of mind, namely, subscription to a relevant moral standard. That is, there are linguistic conventions governing the use of moral terms which are such that, if a speaker asserts a basic moral proposition by uttering a sentence in which a moral term is used literally, the speaker implies, other things being equal, that she subscribes to a relevant moral standard. The argument proceeds in stages. First, I argue that a speaker who makes a typical moral assertion *conversationally* implicates that she subscribes to a relevant moral standard, other things being equal. The argument at this first stage depends on pragmatic considerations about moral discourse. Second, I argue that when a speaker uses moral terms to make a moral assertion, the fact that she implicates that she subscribes to a relevant moral standard does not need to be explained by assuming there is common knowledge of such pragmatic considerations. It appears that it can be explained by reference to linguistic conventions governing the use of the terms. Third, I argue more specifically that it can be explained by the *coloring* of the terms. To show this, I apply the four tests for coloring to moral terms. This overall three-stage argument is not conclusive, for there are different ways of explaining the intuitions elicited by the examples that I will present at various points. I merely claim that the expressivist view I put forth deserves to be taken seriously.⁵⁶

I take it to be common ground, for present purposes, that a person who says something with the intention to make a moral assertion that M sincerity-expresses her belief that M. The explanation for this is simple. To make an assertion is to say something with the intention that one's hearer take one to believe what one says. Accordingly, other things being equal, to assert that M when one does not believe that M is deceptive, assuming one knows what one believes, and it is therefore insincere. The question is whether, other things being equal, a person who makes a moral assertion also expresses in some sense her subscription to a relevant standard.

The chief reason that we teach our children our moral values, and that we want our fellow citizens to share our moral values, is surely that we want them to govern their behavior accordingly. Our aim is not chiefly that people simply agree with us in their judgments. Now, in my view, the

56 I am grateful to Justin D'Arms, Janice Dowell, Don Hubin, Steven Rieber, David Sobel, Sigrún Svavarsdóttir, and David Velleman for help with the arguments in this section.

key action-guiding moral states of mind are states of subscription to moral standards: When we act morally and act well, our behavior is guided by the moral standards to which we subscribe. Hence, I would say, the chief reason we teach our values to our children and want others to share our values is that we want them to subscribe to and endorse our moral standards, so that their behavior will be guided accordingly. Subscription to moral standards, unlike mere moral belief, is partly constituted by relevant intentions, and because of this it brings moral considerations into our planning and decisions about how to act. Indeed, the reason that the currency of a moral code can serve the needs of society is that it involves a widespread subscription to the code within the society's population, and widespread subscription to the code consists in part in a widespread tendency and policy to comply with it. From this perspective, therefore, moral belief is of secondary importance to subscription to moral standards. From this perspective, the chief point of moral teaching is to bring people to have appropriate moral standards, and the point of moral discourse is in part to guide participants in the discourse to morally appropriate action-guiding states of mind – namely, states of subscription to justified standards – so that their actions will be appropriately based in a warranted moral outlook. The point of moral discourse depends, therefore, on our tending to have moral convictions in my sense, moral beliefs that are accompanied by subscription to corresponding standards.

If this account is basically correct, we can see why we would expect a person who expresses a basic moral belief to subscribe to a corresponding standard. As a result of the standard processes of moral teaching, people typically have moral convictions rather than “bare” moral beliefs. That is, if one has a moral belief, one typically also subscribes to a corresponding moral standard. A person who expresses a moral belief therefore typically has the corresponding conviction – and the point of expressing a moral belief is typically, in part, to guide others to share the corresponding conviction. Given all of this, moral conviction is the expected state of mind of a person who expresses a basic moral belief. Because of this, other things being equal, a person who makes a moral assertion conversationally implicates that she subscribes to a corresponding moral standard. This in turn means that it would be misleading of her, other things being equal, to express a basic moral belief if she did not subscribe to a corresponding standard.

This account of the pragmatics of moral assertion would explain why, in Smith's example, it would be *misleading* of me to assert that I ought to give to famine relief if I do not subscribe to a standard that requires

me to give to famine relief. It does not follow, however, that I would be *insincere* to assert that I ought to give to famine relief if I do not subscribe to such a standard, for there need not be any deceit or pretense on my part. Hence, it does not follow that I sincerity-express subscription. We can, however, define a notion of *conversational-expression*. Let us say that a person conversationally-expresses a state of mind in asserting something if and only if, in asserting the thing, she implicates conversationally that she is in that state of mind. Given this definition, we can say that a person who asserts a basic moral proposition *conversationally-expresses* subscription to a corresponding standard. Other things being equal, she implicates conversationally that she subscribes.

Let me use the expression “context of decision” to refer to conversational situations in which the topic at issue is what the speaker should do. The question in contexts of this kind is not merely how to classify various actions under normative properties. The conversation is instead aimed at the speaker’s reaching a decision. Suppose, then, that a speaker who is in such a context says, “Cursing is wrong.” If she understands that the topic of conversation is whether to curse, and if she intends to be participating in the conversation and is aware that the point of moral discourse is to address and assess the moral standards subscribed to by the decision-maker and their bearing on the decision, then she must realize that she will be taken to subscribe to a standard that prohibits cursing. If she does not, then, other things being equal, it would be misleading of her to say that cursing is wrong without canceling the implication that she subscribes to a corresponding standard. It appears therefore that, other things being equal, a person who asserts a basic moral proposition in a context of decision conversationally-expresses her subscription to a corresponding moral standard.

Of course we sometimes discuss hypothetical situations, or decisions made by other people. Even in these cases, however, the point of moral discourse retains its focus on addressing and assessing the moral standards to which we subscribe, for we could find ourselves in situations relevantly like the ones at issue. Hence, I think, it can be misleading even in situations of this kind to make a moral judgment if one does not subscribe to a corresponding moral standard. I would seem hypocritical if, in discussing your cursing, I say “Cursing is wrong,” but have no intention to avoid cursing myself and no tendency to feel negatively about myself if I curse. It thus seems plausible that, other things being equal, a person who asserts a basic moral proposition conversationally-expresses her subscription to a corresponding moral standard.

We can imagine cases in which it would not seem to be misleading to make a moral judgment without subscribing to a corresponding moral standard. Suppose that a person says, "Capital punishment is wrong." Suppose also that she lives in a society that practices and strongly supports capital punishment, that she knows most people around her are strong supporters of capital punishment, and that she knows there is nothing she can do to change the situation. She is now numb to the situation; she has no intention to do anything about the fact that capital punishment is practiced in her society, and she has lost any tendency she once had to respond negatively to those who accept capital punishment. It follows that she does not subscribe – in my sense – to a standard prohibiting capital punishment. Nevertheless, if her state of mind is well understood, it would not be misleading of her to assert that capital punishment is wrong, and there need not be any insincerity in her state of mind. If this is correct, then, in saying capital punishment is wrong, she neither sincerity-expresses nor conversationally-expresses her subscription to a standard that prohibits capital punishment. In situations of this kind, it would not necessarily be insincere or misleading for me to make a moral assertion even if I do not subscribe to a relevant standard. Moreover, in situations of this kind, I do not conversationally implicate that I subscribe to such a standard.⁵⁷

There are also contexts of decision in which "other things" are not "equal." Consider a woman who says, "Cursing is wrong," yet is known to be someone who curses like a sailor. Suppose that a priest is meeting with her and has confronted her about her cursing. The woman says, with a sneer and a chuckle, "I could continue to curse, but of course it would be wrong." She presumably understands that the issue raised by the priest is whether she should continue to curse, but it seems to me that, given what the priest knows of her, he would not take her sincerity here to depend on whether she has any intention to guide her behavior in light of the wrongness of cursing. If the priest is at all worldly, he will not be misled; the woman has not implicated that she subscribes to a standard that prohibits cursing.

I have argued that, other things being equal, a person who asserts a basic moral proposition conversationally-expresses her subscription to a relevant standard. That is, she implicates conversationally that she subscribes to a corresponding moral standard, other things being equal. The argument

57 The example I use here is similar to Michael Stocker's example of the retired politician (1979, 741).

depends on the pragmatics of moral discourse as I see it. Given how participants in moral discourse understand what they are doing, they form certain expectations, and because of this, a person who makes a moral assertion implicates, other things being equal, that she subscribes to a corresponding standard. This explains why it can be misleading to assert a basic moral proposition if one does not subscribe to such a standard.

The argument does not turn on the use of moral terms. Suppose, for example, that Bob says to Alice, who just uttered a curse, “I would not have done that!” Bob might intend thereby to express the belief that it was morally wrong of Alice to curse, and she might realize this. If so, then given the pragmatic features of moral discourse that I have been describing, Bob might also have implied conversationally that he subscribes to a standard that prohibits cursing. Apparently, then, one does not need to use moral terms in order to conversationally-express subscription to a moral standard.⁵⁸ In my view, the familiar moral terms are not the only terms that refer to moral properties. For example, I think that the complex expression “prohibited by the moral code the currency of which in S actually would best contribute to S’s ability to meet its needs” refers (roughly) to the property of moral wrongness. Since hardly anyone accepts the society-centered theory, it would only be in certain special contexts that we would take a person to express a moral belief if she said, “Cursing is prohibited by the moral code the currency of which in this society actually would best contribute to its ability to meet its needs.” In these contexts, a person who said such a thing might conversationally implicate that she subscribes to a relevant moral standard.

It appears, nevertheless, that it would serve the goals of moral discourse if a convention were to develop such that a person asserting a basic moral belief by using moral terms would thereby implicate that she subscribes to a relevant moral standard. If such a convention were to develop, then the connection of moral discourse to action-guiding states of mind would be encoded in the meanings of the terms we use. Under such a convention, if a person were to assert a basic moral belief by using moral terms, we would not need to have any special knowledge of the pragmatics of moral assertion to understand the speaker to be in a state that includes subscription to a standard. As Frege suggested, conventions can develop to govern uses of a term when “it has constantly been used in cases of the same kind.”⁵⁹ Hence, it would not be surprising if it were a feature

58 David Sobel pointed this out, in discussion.

59 Frege 1979c, 141.

of the meaning of “morally wrong” that, other things being equal, when a person says that something is “morally wrong,” she thereby expresses her subscription to a relevant standard. It would not be surprising, that is, if moral terms have a coloring such that when a speaker asserts a basic moral proposition, she thereby implicates that she subscribes to a relevant standard.

To me, it seems plausible that moral terms *do* have such coloring. It would be misleading, other things being equal, for Bob to say to Alice, “Cursing is morally wrong,” if he did not subscribe to standard that prohibits cursing. If Alice understands the point of moral discourse, she will take him to subscribe to such a standard. Now, I do not think that a person needs to have a sophisticated understanding of the pragmatics of moral discourse in order to understand that a person who asserts a basic moral belief in moral terms is implicating subscription to a corresponding moral standard. It seems to me that it is only necessary to understand what is said. If this is correct, then perhaps there are linguistic conventions governing the use of moral terms such that they are standardly used to express subscription to norms. Perhaps, that is, moral terms have a coloring that suits them to express subscription to norms.

To test whether the term “morally wrong” has coloring of this kind, we can apply our four tests for coloring. These are, recall, Frege’s truth test, Grice’s test of detachability, my test of cancelability, and the misuse test. To apply the truth test, consider an amoralist: someone who does not have any moral policies, and so does not subscribe to any moral norms. Suppose she says, “Cursing is morally wrong.” Even though she does not subscribe to a norm that prohibits cursing, we certainly would not conclude on this basis that the proposition she has asserted is false. Second is the test of detachability. It seems to me that inverted commas or “scare-quotes” can be used to decolor terms that are standardly colored. I might say, for example, that such and such behavior would be “unladylike,” indicating by emphasis or gesture that I put the term in scare-quotes. Alternatively, I might speak of “so-called unladylike behavior.”⁶⁰ On this model, the amoralist could have said, “Cursing is ‘morally wrong,’” placing “morally wrong” in inverted commas, or she could have said, “Cursing is so-called morally wrong”; either method would have detached the implication that she has a policy against cursing. Third, consider the test of cancelability. An amoralist could explicitly cancel the implication that she subscribes

60 I was helped with these ideas by discussions with Don Hubin, Steven Rieber, and David Velleman.

to a standard that prohibits cursing by saying, “I agree that cursing is morally wrong, but I certainly have no policy of avoiding cursing.” Finally, we come to the misuse test. It seems to me that, in most contexts, the amoralist would misuse the term “morally wrong” in saying, “Cursing is morally wrong,” since she does not subscribe to any prohibition of cursing. Her use of the term could be challenged, just as a person’s use of the term “Yankee” might be challenged if the person has no contempt for Americans. This is evidence of a misuse. Such challenges can sometimes be answered satisfactorily if the speaker was using the term in scare-quotes. For example, a person who is challenged to explain why she uses the term “Yankee” might respond that she meant to use the term in scare-quotes. Similarly, if the amoralist is challenged to explain why she says that cursing is “morally wrong,” she might respond that she is using the term in scare quotes. Despite this, however, in saying, “Cursing is morally wrong,” the amoralist might well be expressing the belief that cursing is *morally wrong*, which is the same belief that you would express in saying, “Cursing is morally wrong.”⁶¹

Given all of this, it seems plausible to me that the term “morally wrong” has a coloring such that, other things being equal, a person who asserts that an action is “morally wrong,” using the term literally in asserting a moral proposition, *conventionally* implicates that she subscribes to a standard that

61 There is a complication here that I do not want to address in the text. On my view, a person who uses moral terms to assert a basic moral proposition *M* *conventionally* implicates that she subscribes to a corresponding standard. Suppose that an individual decolors the moral terms she uses to express *M* by placing them in scare-quotes. On my view, she still asserts that *M*, for her terms still have their original core meaning. It is part of my view, however, that a person who asserts a moral proposition *M* *conversationally* implicates that she subscribes to a corresponding moral standard, other things being equal. On my view, then, it appears that the person still implicates that she subscribes to a corresponding moral standard, and that therefore she has *not* in fact managed to detach the implication by decoloring her terms. If this is correct, it threatens my thesis that the moral terms are colored in the first place. The solution to this problem is that other things are not equal when one uses a decolored moral term. In decoloring a colored term, the use of which standardly implicates that *p*, one *both* detaches *and* cancels the implication that *p*. To see this, consider the term “heretical,” which I take to be colored. To call a view “heretical” is, I suppose, to implicate conventionally one’s disapproval of those who hold the view. Now suppose that I am engaged in a discussion of the conditions under which a religious view would count as heretical. The various things I say might implicate conversationally that I take heresies seriously, so that I disapprove of people who hold heretical views. My use of the term would conventionally implicate the same thing. However, if I were to decolor the term by placing it in scare-quotes, I would not only detach the conventional implication, but would also cancel the conversational implication. In similar fashion, if I called certain views “so-called heresies,” I would not implicate that I disapprove of those who hold them.

prohibits the action. The speaker *Frege-expresses* her subscription to such a standard. This is explained by the fact that the use of “morally wrong” is governed by expressive linguistic conventions of the kind I have described. This is the heart of the expressivist thesis of the realist-expressivist view I am proposing, as applied to the term “morally wrong.” I believe it is plausible that there are similar expressive conventions governing the use of other moral terms, but I will not attempt to argue the point. The general thesis is that moral terms have coloring such that in asserting a basic moral proposition by uttering a sentence in which a moral term is used literally, a speaker conventionally implicates that she subscribes to a relevant standard. The speaker *Frege-expresses* her subscription.

In summary, then, I have been arguing for the plausibility of the conventional implicature view for cases in which a speaker asserts a basic moral proposition by using moral terms. In such cases, I have been contending, the speaker conventionally implicates, other things being equal, that she subscribes to a corresponding standard. In other cases, I have said, a speaker who expresses a moral belief conversationally implicates that she subscribes to a corresponding standard, other things being equal.

8. THE REALISM IN REALIST-EXPRESSIVISM

Near the beginning of this chapter, I warned that the differences between realist-expressivism and certain other views are quite subtle. Because of this, it is easy to confuse realist-expressivism with, on the one hand, a kind of internalist moral realism, or with, on the other hand, a kind of antirealist-expressivism that exploits deflationism about the meaning of “true.” I now want to underline the differences between these views and realist-expressivism.

Let me begin with the difference between realist-expressivism and familiar internalist varieties of moral realism. As I explained in section 1, internalism in this context is the doctrine that a person who believes that she ought to do something must be motivated to some degree to do it. In Smith’s example, I agree after discussion that I ought to give to famine relief, but soon thereafter I fail to exhibit any motivation to give. Smith says my behavior in this case casts “serious doubt on the sincerity of my claim to have been convinced that it is right to give to famine relief at all.” As I said before, this might suggest that the sincerity of a moral assertion in a context of this kind depends on the speaker’s being appropriately motivated. Even so, and even though the sincerity of a moral assertion

depends on the speaker's believing what she asserts, it would be a mistake to conclude that the speaker's believing what she asserts depends logically on her being appropriately motivated.⁶² This is because the having of the relevant belief and the having of the appropriate motivation might each be necessary to the sincerity of the speaker's assertion without the motivation being a necessary condition of the belief.

As I showed in the [previous section](#), the expressivist half of realist-expressivism can explain Smith's observation in the famine relief example. This does not mean, however, that the expressivist thesis commits realist-expressivism to some standard form of internalism. The expressivist thesis implies simply that a person who *asserts* that she morally ought to do something *implies* that she is motivated to do it. It does not follow that a person who genuinely *believes* that she morally ought to do something must be motivated to do it.⁶³ Therefore, internalism does not follow from the expressivist thesis.

Realist-expressivism does, however, support the thesis that a person who asserts, in so many words, that she "morally ought" to do something must also subscribe to a moral standard that calls on her to do the thing in question – otherwise, unless she somehow cancels or detaches the implication that she subscribes to such a standard, her statement is inappropriate. In short, other things being equal, a person who says that she "morally ought" to do something Frege-expresses her subscription to a relevant standard; she expresses her intention or commitment to do the thing in question. We might call this thesis *discourse internalism*. Realist-expressivism can explain why internalism seems plausible to many moral realists, for the correct view – realist-expressivism – supports this cousin of internalism. Because of this, it is easy to mistake realist-expressivism for a combination of realism with a standard form of internalism.

Let me turn now to the differences between realist-expressivism and antirealist expressivism. It is possible in fact to confuse realist-expressivism with a number of different antirealist views, or at least to think that it is not significantly different from these views. As I have already pointed out, it might seem that realist-expressivism is merely a variant of Blackburn's quasi-realism. Realist-expressivism might also seem very similar to the views of antirealist-expressivists, such as Hare, who agree that there is

62 Smith does not make this mistake.

63 Similarly, a person who calls a dog a "cur" in stating a belief about the dog thereby *expresses* contempt for the dog, but it does not follow that she must actually have contempt in order to have the belief she states.

some element of “description” in moral judgments. Finally, one might think that realist-expressivism merely treats the “thin” moral predicates, such as “wrong,” in the way that standard antirealist-expressivism treats the “thick” moral predicates, such as “honest.” I shall focus on each of these possible confusions in turn.

Let us consider first the issue of whether there is any interesting difference between Blackburn’s quasi-realism and realist-expressivism. The central issue here is whether there is any interesting difference between the realist half of realist-expressivism and the quasi-realist half of Blackburn’s expressivism. Given his deflationist views about the meanings of “true,” “property,” and “belief,” Blackburn would agree with the realist that there are moral truths, moral properties, and moral beliefs. However, Blackburn does not accept the central thesis of moral realism when it is carefully formulated the way I recommend. On my formulation, the central doctrine of moral realism is that moral predicates refer to robust moral properties; they have the same basic semantic characteristics as at least some typical nonmoral descriptive predicates. Blackburn would deny this. He says that, from “inside” moral discourse, there is no discernible difference between what a quasi-realist and a moral realist would be willing to say. However, he says, when one raises “external” questions about morality, such as whether there are moral properties, the quasi-realist holds there are only the attitudes and stances of people.⁶⁴ Presumably, then, he would deny that there are moral properties in addition to these attitudes and stances, although he surely would admit that there are familiar descriptive properties, such as the property of being a perennial, in addition to the attitudes and stances of people. Blackburn therefore denies the central thesis of moral realism when it is formulated the way I recommend. This is the most important difference between Blackburn’s quasi-realism and my realist-expressivism.

Second, there are antirealist-expressivists who agree that there is an element of “description” in moral judgments. Hare says, for example, that moral terms have “descriptive meaning” as well as “evaluative meaning.” In commending things, we apply standards of evaluation to them. If the standard by which a speaker is judging is well known, Hare says, we can infer that she means to attribute certain descriptive characteristics to the thing she is evaluating. Hence, he notes, “If a parson says of a girl that she is a good girl, we can form a shrewd idea, of what description she is; we

64 Blackburn 1997, 173, also 168–169.

may expect her to go to church, for example.” This is “part of what the *parson* means” in calling the girl “good,” says Hare.⁶⁵ Hare puts this point by saying that the “descriptive meaning” of “good” is “secondary to the evaluative.” The realist-expressivist might reverse the order of priority, but it is not clear that this is an interesting difference between realist-expressivism and Hare’s version of expressivism.

Hare ought to have said, however, that although *we* can mean something descriptive in using “good,” the meaning of the *term* is not descriptive.⁶⁶ He agrees that possession of the descriptive characteristics on the basis of which we judge a person to be good is not “entailed” by what we say in calling him good.⁶⁷ In the example of the parson, Hare presumably does not think it is part of the literal meaning of the *sentence* “She is a good girl” that we may expect the girl to go to church, even though he does clearly think that this is part of what the *parson* means to say.⁶⁸ Hare’s considered view, then, should be that at least the thin moral terms have only evaluative meaning. They are used to express moral states of mind, such as prescribing or commending. This brings out the fundamental disagreement between Hare’s version of expressivism and the realist-expressivism that I am proposing, for according to realist-expressivism, moral terms refer to robust moral properties.⁶⁹

Third, on any plausible version of expressivism, thick moral terms, such as “honest,” have the same basic semantic characteristics as at least some nonmoral descriptive predicates. An expressivist like Hare would agree that “honest” refers to some robust nonmoral property, such as, perhaps, the property of being disposed to assert only what one believes.

65 Hare 1952, 146. See also Stevenson 1997, 78.

66 Hare 1952, 148–149.

67 Ibid., 145.

68 Ibid., 109–110, 117–118, and 118–120. In these passages, Hare appears to see that he needs to distinguish between the question, “What do you mean, good?” and the question, “What does ‘good’ mean?” The former asks about speaker’s meaning, the latter about the meaning of the term.

69 Stephen Barker has recently proposed an expressivist view according to which a person making a moral assertion that M both asserts some nonmoral empirical proposition and conventionally implicates that she has a conative or motivational state C-M. I am happy enough with the second part of Barker’s proposal, but according to the first part, a person making a moral judgment does not express a moral belief – instead she expresses an ordinary nonmoral empirical belief. This strikes me as quite implausible. Barker agrees with realist-expressivism that moral terms refer to robust properties, but he has no room in his account for the existence of robust *moral* properties. For this reason, his view qualifies as a kind of antirealist-expressivism rather than as a kind of realist-expressivism. See Barker 2000. I owe this reference to Kent Bach.

Of course, an expressivist would insist that the thick moral predicates are like the thin moral predicates in that they have the semantic role of expressing a characteristic conative state of mind. Nevertheless, the expressivist's view of the semantics of thick moral predicates parallels the realist-expressivist's view of the semantics of thin moral predicates such as "wrong." It might therefore seem that the only new idea in realist-expressivism is the idea that the semantics of the thin moral predicates is not importantly different from the semantics of the thick predicates as it has been conceived in expressivism.

This objection misses the significance of the fact that, according to familiar forms of antirealist-expressivism, the thick moral terms qualify as moral terms only in virtue of the fact that they are used to express a distinctive conative state of mind. Realist-expressivism does not agree with this. It agrees that properties such as the property of being disposed to say what is true are not moral properties. However, it holds that *if* the term "honest" is a moral term, then it does not simply refer to such a property. It refers instead to a robust *moral* property, such as, perhaps, the property of being disposed to assert only what one believes exactly to the extent that being so disposed is virtuous, or the property of being disposed to assert only what one believes exactly to the extent that being so disposed is called for by the ideal moral code of the relevant society. Of course, realist-expressivism also holds that the thin moral predicates refer to robust moral properties. This is independent of the fact that these predicates also have a coloring such that their use standardly expresses the speaker's subscription to a relevant standard.

According to the version of realist-expressivism I am proposing, the proposition that cursing is wrong is a moral proposition. It is a proposition that attributes to cursing the robust property of, roughly, being prohibited by the ideal moral code of the relevant society. If I express belief in this proposition by saying that cursing is "morally wrong," I implicate that I subscribe to a standard prohibiting cursing. This is a realist picture. It is not compatible with an antirealist version of expressivism.

9. CONCLUSION

My goal in this chapter has been to state a version of realist-expressivism, to explain the respects in which it is realist as well as expressivist, and to offer an explication and a partial defense. It seems to me that realist-expressivism has significant theoretical advantages over other views. Let me review some of its strengths.

Realist-expressivism permits moral realists in general and moral naturalists in particular to resist certain standard expressivist arguments against their view by agreeing with expressivists about the expressive quality of moral language. Gibbard claims, for example, that “the special element that makes normative thought and language normative” is that “it involves a kind of endorsement.”⁷⁰ The realist-expressivist can agree that an element of endorsement is involved in normative discourse. The existence of realist-expressivism as a coherent position shows that one need not abandon “descriptivistic analyses” of moral belief in order to acknowledge the element of endorsement. Accordingly, an argument for the positive thesis of expressivism is not an argument for antirealism, because the positive thesis of expressivism can be combined with realism.

Expressivists hold that the “normativity” of moral terminology is captured by the expressive speech-act potential of sentences containing moral terminology. In Gibbard’s words, as I just mentioned, the “element of endorsement” is “what makes normative thought and language normative.” It is not clear what Gibbard means by “normative” in this context. But realist-expressivism acknowledges that there is a difference between the meaning of “wrong” and the meaning of any complex description in nonmoral language that could also be used to pick out the property of wrongness. This is because the term “wrong” has a distinctive coloring in virtue of which it can be used to express subscription to a relevant moral standard. This difference could be described as a difference between the “normativity” of the term *wrong* and the “nonnormativity” of any such complex description in nonmoral language.

Realist-expressivism also acknowledges that there is a difference between the state of mind of a person who believes that, say, cursing is wrong, without subscribing to a prohibition on cursing, and the state of mind of a person who both has this belief and subscribes to such a standard. Indeed, insofar as the point of morality is to guide people’s actions into socially constructive channels, our goal in teaching our children about moral matters is not only to provide them with morally correct beliefs, but also to ensure that they subscribe to corresponding standards so that they will integrate moral goals into their life-planning. We want people to have appropriate moral *convictions*, in my sense of the term, and not merely to have true moral *beliefs*. Now, one could say that moral conviction, in that it involves an action-guiding subscription to a standard, is a “normative” state of mind in a way that bare moral belief is not.

70 Gibbard 1990, 33.

Realist-expressivism can thus acknowledge that there is here a real difference between moral conviction and moral belief that could justify this terminological choice. Hence, realist-expressivists can agree to a point when Gibbard says that normative thought and language derive their normativity from their connection to an “element of endorsement.”

Another theoretical advantage of realist-expressivism is that it gives the realist a way of understanding open-question arguments and claims about the is/ought gap, the naturalistic fallacy, and the like. According to the version of realist-expressivism I have proposed, moral terms have a coloring in virtue of which their meaning, in a wide sense, is not captured by any complex nonmoral descriptive phrase. Similarly, given the coloring of “cur,” its meaning is not captured by “mongrel dog.” According to my version of realist-expressivism, there are expressive linguistic conventions governing the use of moral terms that link them to relevant conative states (such as subscription to moral standards), but there are no such conventions governing the use of the complex nonmoral descriptive phrases that can also be used to refer to moral properties. This is one reason we might detect an “open question,” a “gap,” or a “fallacy,” when a moral realist claims that wrongness is identical to some natural property, such as the property of being prohibited by the ideal moral code of the relevant society. The existence of a gap of this kind, however, is no reason to think that the realist is mistaken to identify moral properties with natural properties.

Finally, as we saw in the [previous section](#), realist-expressivism can capture many of the intuitions that motivate internalist accounts of the connection between moral belief and motivation, and it can do this without being committed to any standard kind of internalism. Realist-expressivism does imply the view I called discourse internalism, which is the thesis that, other things being equal, a person who asserts in so many words that she “morally ought” to do something thereby Frege-expresses her subscription to a corresponding moral standard. In so doing, she expresses her intention or commitment to do the thing in question. But discourse internalism must not be confused with the standard kind of internalism that I have discussed.

In developing realist-expressivism, I have exploited a comparison between moral discourse and “cur discourse.” This is perhaps unfortunate, but it is helpful. Realism holds that there are robust moral properties just as there are robust canine properties, that there are moral truths just as there are truths about canines, and that we have moral knowledge just as we have knowledge about canines. In the version I advocate, realist-expressivism adds that moral terminology gives us a way to express moral beliefs and, at

the same time, to express subscription to corresponding standards. Similarly, we have the term “cur” as well as the term “mongrel dog,” and because of this we have a way to express our beliefs about canines that also expresses contempt for the dogs that we refer to as “curs.” Fundamentally, then, my version of realist-expressivism holds that the semantics of moral terms is similar to the semantics of terms such as “cur.”

The point of moral discourse is at least in part to exploit a social means as well as an articulated means to guide people to form morally appropriate action-guiding states of mind. In order for actions to be based in a moral perspective, moral belief needs typically to be accompanied by subscription to corresponding moral standards. Subscription to moral standards, unlike mere moral belief, is partly constituted by relevant intentions, and because of this it brings moral considerations into our planning and decisions about how to act. Hence, the point of moral discourse depends on our coming to have moral convictions in my sense, and not mere moral beliefs unaccompanied by subscription to corresponding standards. This point is served by the existence of conventions governing the use of our moral terms such that when we use moral terms in expressing our basic moral beliefs we also thereby express our subscription to corresponding moral standards. I have tried to make a case in this chapter that there are conventions of this kind. If I am correct, then moral discourse does not merely express bare moral beliefs; it also expresses the standards we accept and that figure in our decisions about how to act.

REFERENCES

- Austin, J. L. 1962. *How to Do Things with Words*. Cambridge, Mass.: Harvard University Press.
- Ayer, A. J. 1952. *Language, Truth and Logic*, 2nd ed. New York: Dover.
- Bach, Kent. 1999. “The Myth of Conventional Implication.” *Linguistics and Philosophy* 22: 327–366.
- Bach, Kent, and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.
- Barker, Stephen J. 2000. “Is Value Content a Component of Conventional Implicature?” *Analysis* 60: 268–279.
- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- . 1997. “How to Be an Ethical Antirealist.” In Darwall et al. 1997: 167–178. Originally published in: *Midwest Studies in Philosophy* 12 (1988): 361–375.
- Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.

- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In *Moral Knowledge? New Readings in Moral Epistemology*, ed. Walter Sinnott-Armstrong and Mark Timmons, 243–266. New York: Oxford University Press.
- . 1997. "Does Moral Theory Need the Concept of Society?" *Analyse und Kritik* 19: 189–212.
- . 1998. "Morality and Society – The True and the Nasty: Reply to Leist." *Analyse und Kritik* 20: 30–45.
- Darwall, Stephen. 1983. *Impartial Reason*. Ithaca, N.Y.: Cornell University Press.
- Darwall, Stephen, Allan Gibbard, and Peter Railton, eds. 1997. *Moral Discourse and Practice: Some Philosophical Approaches*. New York: Oxford University Press.
- Davis, Steven. 1976. *Philosophy and Language*. Indianapolis, Ind.: Bobbs-Merrill.
- Devitt, Michael. 1997. *Realism and Truth*, 2nd ed. Princeton: Princeton University Press.
- Dreier, James. 1990. "Internalism and Relativism." *Ethics* 101: 6–26.
- Dummett, Michael. 1993. "What Is a Theory of Meaning? (II)." In *The Seas of Language*, 34–93. Oxford: Oxford University Press.
- Field, Hartry. 1994. "Disquotational Truth and Factually Defective Discourse." *Philosophical Review* 103: 405–452.
- Frege, Gottlob. 1979a. "A Brief Survey of My Logical Doctrines." In Frege 1979b: 197–202.
- . 1979b. *Posthumous Writings*, ed. Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach, trans. Peter Long and Roger White. Chicago: University of Chicago Press.
- . 1979c. "Separating a Thought from Its Trappings." In Frege 1979b: 138–149.
- . 1984a. *Collected Papers on Mathematics, Logic, and Philosophy*, ed. Brian McGuinness, trans. M. Black, V. Dudman, P. Geach, H. Kaal, E.-H. W. Kluge, B. McGuinness, and R. H. Stoothoff. Oxford: Basil Blackwell.
- . 1984b. "On Concept and Object." In Frege 1984a: 182–194.
- . 1984c. "On Sense and Meaning." In Frege 1984a: 157–177.
- . 1984d. "Thoughts." In Frege 1984a: 351–372.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Grice, Paul. 1989. *Studies in the Ways of Words*. Cambridge, Mass.: Harvard University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.
- Horwich, Paul. 1993. "Gibbard's Theory of Norms." *Philosophy and Public Affairs* 22: 61–78.
- Jackson, Frank, and Philip Pettit. 1998. "A Problem for Expressivism." *Analysis* 58: 239–251.
- King, Jeffrey C. 1998. "What Is a Philosophical Analysis?" *Philosophical Studies* 90: 155–179.
- Neale, Stephen. 1999. "Coloring and Composition." In *Philosophy and Linguistics*, ed. Kumiko Murasugi and Robert Stainton, 35–82. Boulder, Colo.: Westview Press.
- Sayre-McCord, Geoffrey. 1988. "The Many Moral Realisms." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 1–23. Ithaca, N.Y.: Cornell University Press.

- Searle, John. 1969. *Speech Acts*. Cambridge: Cambridge University Press.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Stevenson, Charles. 1997. "The Emotive Meaning of Ethical Terms." In Darwall et al. 1997: 71–82. Originally published in *Mind* 46 (1937): 14–37.
- Stocker, Michael. 1979. "Desiring the Bad: An Essay in Moral Psychology." *Journal of Philosophy* 76: 738–753.

Milk, Honey, and the Good Life on Moral Twin Earth

Moral naturalism is the view that moral properties, such as rightness and goodness, are in some important sense ‘natural’ properties. Some naturalists have sought to make good on this idea by deploying a kind of semantics proposed by Hilary Putnam and others. Putnam himself suggested this strategy,¹ and Richard Boyd has pursued it.² The kind of semantics proposed by Putnam allows for true synthetic property identity statements, such as that water is H_2O . If this approach can be applied to moral predicates, it therefore opens the door to a kind of ‘synthetic semantic moral naturalism’ according to which, for any moral property, there is a synthetic truth to the effect that the property is identical to a certain natural property. This strategy has come under attack, however, in an argument by Terence Horgan and Mark Timmons known as the “Moral Twin Earth argument.”³ On the one hand, Horgan and Timmons argue, G. E. Moore’s “open question argument” defeated the idea that there are any *analytically* true sentences to the effect that a given moral property is identical to a natural property. On the other hand, they argue, the Moral Twin Earth argument undermines the idea that there are any

I received helpful comments about earlier drafts of this chapter from a number of people. I am especially grateful to Janice Dowell and Jeffrey C. King for extensive written comments and discussion of the issues of semantic naturalism. I am also grateful to Steven Davis, Allan Gibbard, Geoffrey Sayre-McCord, David Sobel, and the members of the Davis Ethics Discussion Group – David Denman, Patrick Findler, Richard Fisher, Lynne Fox, Ron Pritchard, Rick Schubert, and Robert Scofield. A version of this chapter was presented to the Pacific Division of the American Philosophical Association in Los Angeles, March 27, 1998.

1 Putnam 1975b, 244. Richard Boyd cites Putnam 1975a. See Boyd 1988, 199.

2 Boyd 1988.

3 Horgan and Timmons, 1992a, 1992b, 1990–1991.

synthetic truths to the effect that a given moral property is identical to a natural property. If they are correct, naturalism is in trouble.⁴

Are they correct? I will argue that they are not. A close look shows that the Moral Twin Earth argument poses no threat to moral naturalism. To say this is not to say that synthetic moral naturalism is without difficulties. I believe that in fact it has difficulty accounting for the normativity of moral claims. My goal in this chapter, however, is the modest one of making Moral Twin Earth safe for naturalists.

1. SYNTHETIC MORAL NATURALISM

I need to begin by explaining the basic idea of Putnam's semantics. Putnam holds that the extension of a term such as "water" is determined by two factors. The first is the referential intention with which we use the term. Perhaps we use the term "water" with the intention to refer to the liquid that has the underlying nature of the local samples that we would point to in explaining what "water" means. Putnam says we intend to refer to anything that bears the "same liquid relation" or "same_L relation" to the relevant local stuff.⁵ The second factor in the case of "water," given the content of our referential intention, is the nature of this local stuff. This local stuff is H₂O. These two factors determine that the extension of our term "water" is, Putnam says, "the set of all wholes consisting of H₂O molecules, or something like that."⁶ We can say that the term *refers* to this extension. What the local samples have in common, in virtue of which they are related by the same_L relation, is that they have the property of consisting in H₂O.⁷ I will say that our referential intentions together with this fact about the local samples determine that the term "water" *expresses* the property of being H₂O.⁸

Putnam says that similar accounts could be given for "many other kinds of words," including "the great majority of all nouns, and . . . other parts of speech as well."⁹ Consider, for example, artifact terms, such as "soda pop," and "functional terms," such as "milk." As Putnam says, the same_L relation that plays a role in determining the referent of "water" obtains between samples that agree "in important physical properties," but, he

4 Horgan and Timmons 1992a, esp. 153–157.

5 Putnam 1975b, 225.

6 Ibid., 224.

7 Ibid., 239.

8 Ibid., 229–238.

9 Ibid., 242.

stresses, the notion of importance is “interest relative.” Although “structurally important properties” are normally the important ones, in some cases other properties are decisive.¹⁰ The term “milk” illustrates the point, for cow milk and goat milk and human milk have different chemical compositions. Clearly, then, the “physical properties” that determine whether two samples bear the same_L relation to each other need not be chemical in nature. Milk is in fact a liquid that is produced by the mammary glands of female mammals that have recently given birth, in order to nourish their young.¹¹ It is presumably this ‘functional’ and ‘genetic’ property that, on Putnam’s account, would be relevant to determining whether a glass of a white liquid bears the same_L relation to the milk in my fridge.¹² The term “soda pop” also illustrates the point that genetic properties can be decisive. Putnam would say that, when we use the term “soda pop,” “we intend to refer to whatever has the same *nature* as the normal examples of the local [pop] in the actual world.”¹³ But, of course, pop is in fact an artifact. This property of the stuff, along with other facts about the actual nature of the stuff and the content of our intentions, determines the extension of the term “soda pop.” The term “soda pop” illustrates an additional point. It shows that the properties that determine whether samples of a liquid bear the same_L relation to each other need not be ones that play an explanatory role in any science. Soda pop is not a natural kind and “soda pop” is not a natural kind term. Putnam’s semantics is not restricted to “natural kind terms.”

In general, the referent of a term *t* is determined by the conspiracy between the actual nature of the stuff in the samples we use in explaining *t*, on the one hand, and our referential intentions, on the other hand. In simple cases, we intend to use *t* to refer to whatever has the relevantly same nature as these samples. But the examples of functional terms and artifact terms show I think that the theory needs to be developed to allow it to take into account all relevant aspects of the content of our referential intentions. Clearly I have given only a sketch of a theory. As I will explain, a fully developed theory of this kind would need to be amended to take account of the possibility of various kinds of mistakes. But instead of worrying about the details, I will simply say that, for Putnam, there is a

10 Ibid., 239, 241.

11 *American Heritage Dictionary of the English Language*, 3rd ed. (Boston: Houghton Mifflin, 1992), s.v. “milk.”

12 Putnam does not discuss an example sufficiently similar to “milk” for me to be confident how he would treat “milk.” I return to the example below.

13 Putnam 1975b, 243. In this passage, Putnam is discussing “pencil,” not “soda pop.”

complex relation, the “Putnam-relation,” and the referent of a term is the set of “objects” to which the term is “Putnam-related.”¹⁴

Given our limited purposes, this would be sufficient explanation of the kind of semantics that is at issue if it weren’t for the fact that Boyd, whose views are explicitly addressed by Horgan and Timmons, formulates his position in terms of a somewhat different “causal theory of reference.” According to Boyd, the causal relations that are relevant to reference are those “which are involved in the reliable regulation of belief.” He proposes that “a term *t* refers to a kind . . . *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k*” – he asks us to ignore his blurring of the use-mention distinction in this formulation. When there exist such mechanisms, Boyd writes, “we may think of the properties of *k* as [causally] regulating the use of *t*.”¹⁵ Boyd presumably thinks, then, that our practices in using the term “water” are such that the beliefs that we would express using sentences in which the term appears express approximately true propositions about wholes that consist of H₂O, and, over time, our “water” beliefs will tend to become more accurate. We could schematize the view by saying that, for Boyd, there is a complex causal relation, the “Boyd-relation,” and the referent of a term like “water” is the “kind” to which the term is “Boyd-related.”

Boyd’s causal account and Putnam’s intentional account apparently could come apart in certain circumstances. For example, in using the term “honey,” we might intend to refer to the kind of stuff that has the nature of the honey to which we have been exposed, and it might be that the honey to which we have been exposed has all been of a given nature. But despite this, we might have astonishingly inaccurate and even crazy mythological beliefs about honey, such that, in fact, there is no tendency for our ‘honey beliefs’ to get any more accurate. In this case, Putnam’s theory says that “honey” refers to a substance produced by bees, but it appears that Boyd’s theory implausibly says otherwise. For in this imaginary case, the term “honey” is Putnam-related to this substance, but it appears to be Boyd-related to nothing. Given this objection to Boyd’s proposal, I will

14 Strictly speaking, the key relation for Putnam takes a term and a population of speakers who use the term to the term’s referent. It might have other relata as well. This is one of the details I want to finesse.

15 Boyd 1988, 195.

focus attention in what follows mainly on Putnam's account rather than Boyd's.

How do these accounts apply to moral terms? Boyd writes that a moral realist should argue that "Moral goodness . . . , to some extent, regulate[s] the use of the word 'good' in moral reasoning."¹⁶ That is, moral goodness stands in the Boyd-relation to "good." Moreover, our best moral theory gives us our best account of what moral goodness actually is and so it gives our best account of the referent of "good."¹⁷ That is, moral theory plays the role in establishing the referent of moral terms that chemistry plays in establishing the referent of terms like "water." Moral theory gives us our best account of the relevant nature that is shared by standard samples of things that qualify as "morally good," "morally wrong," and the like.

Putnam's view is less straightforward. To apply it to moral terms, we need to know the content of the referential intention with which we use the terms. In using a moral term like "right" or "wrong" or "good," perhaps we intend to refer to "whatever has the same *nature* as the normal examples of the [right or wrong actions or morally good things] in the actual world."¹⁸ Even if so, however, as we saw, different facts about the nature of things are relevant in different cases. The "same liquid" relation obviously is not involved in determining the referent of moral terms. Geoffrey Sayre-McCord suggests that what is relevant is that things in the extension of a moral term be of the relevantly "same moral kind."¹⁹ We can incorporate this suggestion into Putnam's account by assuming provisionally that, in using moral terms, we intend to refer to whatever bears the "moral-same-kind relation" to the examples in the real world that we point to in explaining what such terms mean. The actual nature of these examples conspires with the actual content of our referential intentions to determine what the extension of our moral terms actually is, and our best moral theory gives us our best account of the morally relevant properties shared by these examples. If we ignore for the moment the possibility of mistakes in usage, we can say, for example, that the actions that are actually in the extension of the term "wrong" are those that are

16 Ibid., 201.

17 Ibid., 205. Of course, on standard assumptions, our best theory might be false. The referent of "good" is determined by what goodness actually is, not by our best theory.

18 Putnam 1975b, 243. In this passage, he is discussing "pencil."

19 Sayre-McCord 1997.

of the same moral kind as actions that we would use as examples of wrong actions in teaching what “wrong” means. The property shared by these actions, in virtue of which they are related by the same-moral-kind relation, is the property of wrongness.²⁰

2. THE MORAL TWIN EARTH ARGUMENT

Horgan and Timmons invite us to imagine a variation of Putnam’s famous Twin Earth thought experiment.²¹ Twin Earth is a planet that is “much like” Earth with “exactly similar geography and natural surroundings.” The “Twin English” spoken on Twin Earth is much like the English spoken on Earth. English speakers on Twin Earth use “twin moral” terms that are orthographically just like our moral terms, so they have a term “right” that they use in evaluating actions. There is, in fact, “significant agreement” between the Earthlings’ moral beliefs and the Twin Earthlings’ twin-moral beliefs in the sense that most actions that the Earthlings believe to be wrong the Twin Earthlings believe to be twin-wrong, and so on. In addition, Earthlings and Twin Earthlings agree in their morally relevant nonmoral beliefs. If English-speaking explorers from Earth ever visited Twin Earth, they would be “strongly inclined to translate” the Twin Earth term “right,” and each of their other twin-moral terms, by the orthographically identical English term. For the Earthlings would see that the Twin Earthlings are normally disposed to do what they judge to be twin-right; they normally take considerations about what is twin-right to be quite important in deciding what to do; they criticize others for failing to do what is twin-right. Let us suppose, however, that, on Earth, uses of the term “right” and other moral terms are “causally regulated” by consequentialist properties, the nature of which is captured by a consequentialist moral theory T_C . On Twin Earth, uses of the corresponding Twin English terms are causally regulated by properties whose nature is

20 Our best moral theory gives us our best account of this property. The *correct* theory, if there is one, gives the *correct* account of this property. If we come to think that there is no such property, there are the familiar options for the semantics of moral terms. One could adopt an error theory, along the lines proposed by Mackie (1977), according to which familiar simple moral propositions are false, or a noncognitivist or nondescriptivist theory, along the lines proposed by Allan Gibbard (1990) and others, according to which the surface grammar of moral claims is misleading and they do not express propositions at all.

21 The example is used in three papers. I primarily work from the version in Horgan and Timmons 1992a, 244–246. See also Horgan and Timmons 1992b, 163–166; and 1990–1991, 457–460.

captured by a deontological moral theory, T_D . That is, the Twin Earthlings' uses of their twin-moral terms are causally regulated by properties that are distinct from those that regulate the Earthlings' uses of the orthographically identical moral terms.

Given these assumptions, say Horgan and Timmons, a causal theory of reference along the lines proposed by Boyd would imply that corresponding Earthling and Twin Earthling terms "right," "wrong," and the like "differ in meaning." For it would imply that the terms used by the Earthlings express the consequentialist properties that regulate their use and that the terms used by the Twin Earthlings express the distinct deontological properties that regulate their use. If this implication of the theory were correct, however, say Horgan and Timmons, it ought to seem intuitively plausible to us that the terms differ in meaning. But this does not seem plausible, and so Boyd's semantics must be incorrect. Let us pause to consider these ideas.

In his original Twin Earth thought experiment, Putnam stipulated that although the stuff called "water" on Earth is actually H_2O , the stuff called "water" on Twin Earth is a different liquid, XYZ. Almost everyone has the intuition in this case that the Earthling term "water" differs in meaning from the Twin Earthling term and that any apparent disagreements between Earthlings and Twin Earthlings about whether a given sample of liquid is "water" would be merely verbal. In Horgan and Timmons's Moral Twin Earth scenario, however, our intuitions do not line up with the apparent implications of Boyd's theory in the corresponding way. Suppose, for example, that there is a situation in which the Twin Earthlings judge that it would be "wrong" to lie but the Earthlings judge that it would *not* be "wrong" to lie. According to Boyd's theory, say Horgan and Timmons, their disagreement might merely be verbal, for, according to the theory, the lie could have the property that is expressed by the Twin Earthlings' term "wrong" yet lack the property that is expressed by the Earthlings' term "wrong." In principle the Earthlings and Twin Earthlings could come to see this. Accordingly, we should have the intuition that their disagreement might be merely verbal. Yet, Horgan and Timmons report, and I tend to agree with them, we do not have this intuition. Rather, in the situation I imagined, it seems to us that the Earthlings and Twin Earthlings disagree about whether it would be *wrong to lie*. This intuition implies that the terms on the two planets "do *not* differ in meaning or reference, and hence that any apparent moral disagreements that might arise between Earthlings and Twin Earthlings would be *genuine*

disagreements,” differences in moral belief.²² Horgan and Timmons say that it would seem “wrongheaded” to view these differences as reflecting nothing but differences in the meanings of their terms.²³

One might wonder about the significance of our intuitions in thought experiments of this kind. Horgan and Timmons hold that our intuition that the Earthlings and Twin Earthlings have genuine moral disagreements shows that, contrary to what Boyd thinks, it is false that moral terms express the natural properties that uniquely causally regulate their use.²⁴ For, they hold, our “semantic intuitions” can provide “powerful *empirical* evidence” for or against semantic claims, such as Boyd’s claim about the semantics of moral terms. They concede that such evidence is “defeasible,” but our intuitions about such matters reflect our linguistic competence, which is a competence with the semantic norms that ground the meanings of terms. Hence, the fact that we are intuitively convinced that the Twin Earthlings and the Earthlings have genuine moral disagreements, and not merely verbal disagreements, is strong evidence that they *do* have genuine disagreements. And this is evidence that Boyd’s account of moral terms is mistaken.²⁵

So far we have only an argument against Boyd’s account. But Horgan and Timmons think that our intuitions in Moral Twin Earth scenarios provide evidence against the new naturalistic program more generally. The program is to defend some version of “synthetic semantic naturalism,” the doctrine that moral terms “have synthetic naturalistic definitions” in virtue of the fact that they express natural properties. Horgan and Timmons claim that *any* theory on which moral terms express natural properties can be undermined by means of a version of the Moral Twin Earth argument. For, on *any* such theory, there will be some relation *R* such that terms are held to express the natural properties to which they are *R*-related. We could then construct a Twin Earth example in which the moral terms on Earth are *R*-related to natural properties that satisfy a normative moral theory *T*, and the twin-moral terms on Twin Earth are *R*-related to distinct properties that satisfy a different moral theory *T*'. The argument would then proceed as before. In any Moral Twin Earth scenario, the semantic theory at issue would imply that moral and twin-moral terms differ in meaning and that the Earthlings and the Twin

22 Horgan and Timmons 1992b, 165.

23 Horgan and Timmons 1992a, 248.

24 Ibid., 248, 243.

25 Ibid., 257 n. 37; Horgan and Timmons 1992b, 166.

Earthlings do not have genuine moral disagreements. We would find these implications of the theory to be intuitively implausible, and this, Horgan and Timmons claim, is strong evidence against the theory. The fact that we would have similar intuitions throughout the entire range of Moral Twin Earth scenarios is strong evidence against the program of synthetic moral naturalism quite generally.²⁶

What should we make of this argument? A naturalist might, of course, deny the intuition on which the argument depends, or else deny that it is very strong evidence against synthetic moral naturalism.²⁷ But I think that the intuition that the Earthlings and Twin Earthlings have substantive moral disagreements, and not merely verbal disagreements, is robust and widely shared. For this reason, naturalism must try to accommodate the intuition in some way. I will argue that the intuition can indeed be accommodated by being interpreted in such a way that it does not undermine synthetic naturalism. This is my first reply to the argument. I will also argue that a plausible form of synthetic moral naturalism can be devised according to which Earthling moral terms and corresponding Twin Earthling twin-moral terms do express the same properties. If I am correct, a coherent Moral Twin Earth scenario cannot be devised that creates difficulty for this form of synthetic naturalism. This is my second and more important reply. But let me begin with a point of clarification.

As we saw, Horgan and Timmons mean to attack a view they call “synthetic semantic naturalism,” the doctrine that “Fundamental moral terms like ‘good’ have synthetic naturalistic definitions.”²⁸ By a synthetic naturalistic definition, they mean a statement, such as “Water is H₂O,” that gives the “real nature or essence of the entity, property, or kind” in question.²⁹ Synthetic moral naturalism proposes that the property expressed by a moral predicate is identical to the property expressed by a corresponding nonmoral predicate couched in naturalistic vocabulary, one that plainly expresses a natural property. Notice that this is at least in part a *semantic* doctrine, while moral naturalism as such is rather the *metaphysical* doctrine that, quite simply, moral properties are natural properties. A moral naturalist might therefore argue simply that moral properties have the meta-property of being natural properties without attempting to defend

26 Horgan and Timmons 1992b, 167; 1992a, 248.

27 This line of thought has been discussed by Gampel (1997).

28 Horgan and Timmons 1992b, 167.

29 Ibid., 157.

the existence of synthetic naturalistic definitions.³⁰ Of course, such a naturalist would need to provide an account of what it is for a property to be a ‘natural’ property as well as a semantics for moral terms. These are things that any naturalist must do. The important point is that Moral Twin Earth scenarios could not be used to undermine a moral naturalism of this kind, assuming that such a theory did not imply anything about the circumstances under which a Twin Earth twin-moral term would express a property different from the property expressed by the corresponding Earthling moral term. In what follows, then, we restrict attention to forms of synthetic semantic moral naturalism.

3. THE FIRST REPLY: REINTERPRETING OUR INTUITIONS

I shall argue that synthetic moral naturalism can accommodate the intuition we have in Moral Twin Earth scenarios that the Earthlings and Twin Earthlings disagree morally. Consider again a situation in which the Earthlings say in English that it would *not* be “wrong” to lie but the Twin Earthlings say in Twin English that it *would* be “wrong” to lie. The Earthlings think that lying in this situation would not be *wrong*, while the Twin Earthlings think it would be *twin-wrong*. The Moral Twin Earth argument depends on the premise that if the Earthlings and the Twin Earthlings do in fact disagree morally in such a situation, then the Earthling term “wrong” must express the same property as is expressed by the Twin Earthling term “wrong.” I will argue to the contrary that *even if* the Earthling term “wrong” expresses a property that is distinct from the property expressed by the Twin Earthling term “wrong” – which, of course, is what Horgan and Timmons think that any form of synthetic moral naturalism would be committed to saying in a relevant Moral Twin Earth thought experiment – it still might be the case that the Earthlings and Twin Earthlings disagree morally in the situation we have imagined.

For the sake of argument, let me stipulate that in a new Moral Twin Earth scenario, which I will call the “Master” scenario, Earthling moral terms do in fact express natural properties different from those expressed by the corresponding twin-moral terms. For example, I will assume the Earthling term “wrong” expresses the property “fails to maximize the good,” but the Twin Earthling term “wrong” expresses the property

30 Nicholas L. Sturgeon appears to favor a nonreductionist naturalism (see Sturgeon 1984).

“fails to comply with the rules specified in deontological theory T_D .” Since the scenario is otherwise just like the scenario devised by Horgan and Timmons, we should still have the intuition that the Twin Earthlings and the Earthlings disagree morally. And according to the key premise in Horgan and Timmons’s argument, our intuition that the Earthlings disagree morally with the Twin Earthlings conflicts with the stipulation that the Earthling term “wrong” and Twin Earthling term “wrong” express different properties.

Note, however, that in the Master scenario there is significant overlap between the extensions of corresponding moral and twin-moral terms. Horgan and Timmons stipulate that in Moral Twin Earth scenarios there is “significant agreement” between the Earthlings’ moral beliefs and the Twin Earthlings’ twin-moral beliefs. And it should be clear that there must be such agreement given that Earthling moral terms express consequentialist properties and that Twin Earthling twin-moral terms express deontological properties, for in worlds that are as similar to the actual world as Twin Earth is stipulated to be, consequentialists and deontologists agree in very many cases about which actions are “right” and “wrong.” Hence, by construction of the Twin Earth scenario, most actions that the Earthlings believe to be wrong the Twin Earthlings believe to be twin-wrong, and so on. Moreover, twin-moral judgments play the same role in Twin Earth life that moral judgments play in the life of Earthlings. Twin-moral judgments are normative in just the way that Earthlings’ moral judgments are normative, and Twin Earthlings tend to make negative twin-moral appraisals using “wrong” in circumstances similar to those in which Earthlings make negative moral appraisals using “wrong.” Given these facts about the Master scenario, despite the fact that corresponding moral and twin-moral terms do not express the same property, I now want to argue that moral terms might be the best *translation* for the corresponding twin-moral terms. The best translation of the twin-moral term “wrong” into Earthling English might be the term “wrong.” The explorers from Earth who were inclined to translate the twin-moral term “wrong” by the English word “wrong” might have been correct.

To understand this point, it needs to be understood that translation is more like trying to find someone who looks enough like you to pass muster in a police lineup than it is like trying to find your identical twin. Many people lack identical twins, of course, and many words lack precise synonyms in certain other languages. So it is possible that the English term “wrong” is the best translation into English of the Twin English term “wrong” even though, by stipulation, the terms express

different properties. If this is so, then even though corresponding moral and twin-moral terms express different properties and therefore have different ‘meanings’ in the philosophically preferred sense of the term, there is also a sense in which they might have the same ‘meaning.’ Given this, we can make sense of our intuition that Earthlings and Twin Earthlings disagree morally even if corresponding moral and twin-moral terms express distinct properties. To see this, consider again the situation we imagined in which the Twin Earthlings judge that lying would be twin-wrong but the Earthlings judge that lying would not be wrong. If Earthling “wrong” is the best translation for the Twin Earthling “wrong,” then the Earthlings and Twin Earthlings are plausibly taken to disagree, for the Twin Earthlings can best be understood to judge that lying would be wrong, and the Earthlings do not agree with this.

One might object that, by hypothesis, we have a better translation for the Twin Earthling term “wrong,” namely, “violates one of the rules specified in T_D .” For we imagined that the English term “wrong” expresses the property of failing to maximize the general good, and that the Twin English term “wrong” expresses the distinct property of violating one of the rules specified in T_D . This proposal ignores, however, that translation needs to pay attention to conventions governing contexts of use as well as to the properties that are expressed by terms. It is for this reason that in many contexts it would be misleading to translate a scandalous French term for a bodily function with an innocuous English description of that bodily function even if the term and the description pick out the same bodily function. By construction of the Twin Earth scenarios, Earthling moral terms and Twin Earthling twin-moral terms play the same role in people’s lives on the two planets. The term “wrong” in Twin English is like the English term “wrong” in that it is standardly and conventionally used to express negative appraisals of actions. But the English expression “violates one of the rules specified in T_D ” is not standardly and conventionally used in this way. Because of this, it would be misleading to use the latter to translate the former. Moreover, people on Earth do not believe that it is always wrong to violate one of the rules specified in T_D , so they would not always be inclined to infer that a negative evaluation is intended when an action is described as being in violation of one of the rules in T_D . This is another reason why the proposed translation would be misleading.

Given that, by stipulation, corresponding moral and twin-moral terms express different properties, the Earthlings and the Twin Earthlings would not actually contradict each other in a situation where the Twin Earthlings

said that lying would be “wrong” and the Earthlings denied that it would be “wrong.” We can presume them to know that they would not contradict each other, for we can presume them to know that an action could have both the property of not being wrong and the property of being twin-wrong. Even so, they still disagree about something morally substantive, for they disagree about whether to lie in the imagined situation. They also would disagree about whether to urge people in similar situations not to lie, about whether to object to lying in such situations, about whether to teach their children not to lie in such situations, and so on. They do therefore disagree about something that is morally substantive. Important matters can be at stake here, including whether to avoid actions that are wrong or whether instead to avoid actions that are twin-wrong, whether to teach children the moral code or instead to teach the twin-moral code, and so on. Our intuition that the Earthlings and the Twin Earthlings have something morally substantive to disagree about is therefore true, even if corresponding moral and twin-moral terms refer to different properties.

I conclude therefore that, even given the stipulation that corresponding moral and twin-moral terms express different properties, we might still have the intuition, and it might be correct, that the Earthlings and Twin Earthlings disagree about something substantive and that their terms have roughly the same meaning, in that they are inter-translatable. This means that our intuitions in the original Twin Earth scenarios that the Earthlings and Twin Earthlings disagree are not incompatible with the thesis that their terms express different properties. Our intuitions do not undermine the thesis that moral terms refer to natural properties, nor do they undermine synthetic semantic naturalism.

Putnam’s original Twin Earth argument rested on two intuitions: First, that the Earthling term “water” differs in meaning from the Twin Earthling term, given that the stuff on Earth is H_2O , while the stuff on Twin Earth is XYZ, and second, that any apparent disagreements between Earthlings and Twin Earthlings about whether a given sample of liquid is “water” would be merely verbal. In Horgan and Timmons’s scenario, however, we have the intuition that the relevant English and the Twin English terms have the *same* meaning and that there are *real* moral disagreements between the Earthlings and the Twin Earthlings. Because we have these intuitions, the Moral Twin Earth scenario cannot be used to *support* a Putnamian account of the semantics of moral terms in the way that Putnam used our intuitions in the original example to support his account of the semantics of “water.” What I have argued, however, is

that our intuitions in the Moral Twin Earth scenario do not *undermine* the apparent implication of Putnam's theory that the English and Twin English terms have different meanings in the preferred philosophical sense. In an *ordinary* sense of the term, terms have the same "meaning" if they are inter-translatable. Moreover, in an *ordinary* sense, people have a substantive disagreement if they disagree about what to do. In Horgan and Timmons's example, I have argued that it would be misleading *not* to translate the Twin English term "wrong" by the English term "wrong" even if the terms express different properties. And there are disagreements between the Earthlings and the Twin Earthlings about what to do. These facts explain why it is that our intuitions do not undermine the thesis that corresponding English and Twin English moral terms express different properties in the Moral Twin Earth scenario.

The mistake made by Horgan and Timmons, then, was to take our intuitions in the Moral Twin Earth scenario to be strong evidence that corresponding English and Twin English moral terms express the same property. I have argued that there is an interpretation of our intuitions on which they are compatible with the thesis that corresponding terms in their languages express distinct properties. Our intuitions can therefore be accommodated by semantic theories of the kind that are in question.

4. THE SECOND REPLY: INTRODUCING A THEORY OF ERROR

The first reply to Horgan and Timmons's argument will seem unsatisfying to many philosophers, including, presumably, Horgan and Timmons. For as far as the first reply goes, even if the best English translation of what the Twin Earthlings say in the lying example is that lying would be "wrong," it is nevertheless true in the example that, in saying that lying would be twin-wrong, the Twin Earthlings attribute to lying a property that the Earthlings do not deny it to have when they say that lying would not be wrong. As far as the first reply goes, the English term "wrong" has a different meaning in the philosophically preferred sense from the corresponding Twin English term. It may seem to many philosophers that any theory with this implication is unacceptable. For it may seem intuitively clear that the Earthlings disagree with the Twin Earthlings precisely about the truth of the proposition about lying expressed by the Twin Earthling sentence, "Lying is wrong." It may seem clear that the Earthlings are denying exactly what the Twin Earthlings affirm when they say lying would be "wrong."

I believe that synthetic moral naturalism can accommodate this intuition. In fact, I think that a Putnamian semantics can accommodate the intuition by deploying strategies some of which Putnam himself uses to explain the possibility of errors of various kinds in nonmoral examples, together with a distinction that Putnam draws between speakers' referential intentions and their interests. The resulting semantics implies that corresponding Earthling and Twin Earthling moral and twin-moral terms in Twin Earth scenarios express the same property if they express any property at all. It therefore would imply that in the lying example either the Earthlings or the Twin Earthlings are mistaken. It would also imply that there is no coherent Moral Twin Earth scenario of the sort described by Horgan and Timmons. That is, in any coherent scenario of the kind described by Horgan and Timmons, the semantics would imply that corresponding Earthling and Twin Earthling moral and twin-moral terms express the same property. This is the second reply to the Moral Twin Earth argument. In discussing it, we obviously need to leave behind the special stipulation of the Master scenario that Earthling and Twin Earthling moral and twin-moral terms express different properties.

I should stress that my goal is simply to defend synthetic moral naturalism against the Moral Twin Earth argument. I do not aim to defend it or Putnamian semantics against all worries that could be raised. Accordingly, I aim merely to show that a plausible articulation of Putnamian semantics could be used in explicating synthetic moral naturalism.

Putnam recognizes that his basic semantics needs to be amended to allow for mistakes of various kinds. Suppose, for instance, that I point to a glass that contains a clear liquid and say, "This liquid is called water." Putnam says that my "ostensive definition" of "water" in this example presupposes that the liquid in the glass is the same kind of liquid as most of the stuff that I and most speakers in my linguistic community have been calling "water." If this presupposition is false because, say, the glass actually contains gin rather than water, then, Putnam says, "I do not intend my ostensive definition to be accepted." In general, Putnam says, ostensive definition has certain empirical presuppositions, and when these presuppositions are false, "a series of, so to speak, 'fallback conditions' becomes activated."³¹ Putnam does not say much about these fallback conditions. But he does say that, for natural kind terms, "if there is a

31 Putnam 1975b, 225.

hidden structure, then generally it determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds.” He adds that “the local water, or whatever, may have two or more hidden structures – or so many that ‘hidden structure’ becomes irrelevant, and superficial characteristics become the decisive ones.”³²

We can see how the idea of fallback conditions might work if we compare the semantics of the term “milk” with the semantics of the term “jade.” Both jadeite and nephrite fall in the extension of the term “jade.” There is no hidden structure shared by all samples of jade. All samples do have the disjunctive property of being either jadeite or nephrite, but this of course does not mean that there is a common structure in any relevant sense.³³ Now, just as there are different kinds of jade, there are various kinds of milk, each with a somewhat different chemical composition. Cow milk is chemically different from goat milk. But, even though there is not a chemical composition shared by all milk, there is a “hidden structure” shared by all milk, for all kinds of milk have the functional and genetic property of being, roughly, a liquid that is produced by the mammary glands of a female mammal that has recently given birth, in order to nourish its young. So Putnam’s semantics would presumably account for the semantics of “milk” by reference to the referential intention to refer to whatever has the relevantly same nature as the local samples, given that milk has a shared nature at the level of functional and genetic properties.

Consider another kind of example. Suppose that we have been using narwhale tusks as samples of unicorn horns in explaining the meaning of “unicorn.” We have pointed at narwhale tusks and said such things as, “The animal that this horn came from is called a unicorn.” Putnam’s theory clearly ought not to produce the result that narwhales are unicorns, and it can avoid this result if we make a plausible assumption about the empirical presuppositions of our ostensive definition of “unicorn.” Just as we use “water” with the intention to refer to a liquid, it is plausible that we use the term “unicorn” with the intention to refer to a four-footed single-horned animal that lives on land. If so, then our ostensive definition presupposes that the tusks we are pointing to once adorned such an animal. This presupposition is of course false, and so, Putnam can say, we do not intend our definition to be accepted. Narwhales are not in the extension of “unicorn.” The extension of “unicorn” is in fact empty.

32 Ibid., 241.

33 Ibid.

A full development of a Putnamian semantics would have to explain how to determine what our referential intentions are for any given term. Presumably we can infer the content of the relevant intentions from our linguistic intuitions, given the Putnamian theory. For instance, if the intention with which we use “unicorn” had been simply to refer to whatever animal used to be adorned by these horns, then narwhales would have been in the extension of “unicorn” according to the theory. But since our intuition is that narwhales are not in the extension of “unicorn,” we can infer that this was not our referential intention. We can infer that our intention, whatever its precise content, is such that, according to the theory, narwhales are not in the extension of “unicorn.” Similarly, in the original Twin Earth scenario, our intuition is that the XYZ on Twin Earth is not water. This intuition is explained by Putnam’s semantics on the assumption that the intention with which we actually use the term “water” on Earth is to refer to whatever liquid has the same basic nature as the stuff around us.³⁴

We are seeing that a plausible naturalistic semantic theory would need to incorporate a theory of error. In Putnam’s view, for example, the referent of a term, such as “water,” is determined by the actual nature of the stuff that is in the local samples, not by our beliefs about the nature of the stuff. If our beliefs about the nature of the stuff are mistaken, we can make mistakes in using the term. A plausible theory obviously must allow for such mistakes. Similarly, in the “unicorn” example, we see that certain presuppositions of our usage can be false in a way that overrides the fact that we actually apply the term to narwhale tusks and that leads the theory to the result that the extension of “unicorn” is empty. In the moral case, a properly developed Putnamian semantics would have to allow in similar ways for the possibility of mistakes. It is the moral facts together with our referential intentions that determine the extension of “wrong.” If our moral views are mistaken, then we might make mistakes in calling actions “right” or “wrong.” Or if there are certain false presuppositions of

34 I am making an epistemic proposal about how we can identify the content of the relevant referential intentions. The proposal is that we *infer* this from the theory, together with our semantic intuitions. This procedure is not question-begging, for if the content we need to suppose in order to explain our intuitions in light of the theory is implausible, then, of course, the implausibility counts against the theory. It is not implausible, however, that we use “unicorn” with the intention to refer to a four-footed single-horned animal that lives on land. The procedure does not make the theory circular, for regardless of how we *identify* the content of the relevant intentions, their *actual* content is of course what matters to the theory.

our usage, then fallback considerations would come into play. The theory needs to allow for such mistakes as well.³⁵

Let me now return to the Moral Twin Earth example in order to reconstruct Horgan and Timmons's argument. I am assuming that we share the intuition that the English and Twin English terms "wrong" have the same meaning and that the Earthlings and Twin Earthlings disagree in the lying case. The issue is whether a Putnamian semantics can explain the intuition. Recall that in Putnam's semantics, the referent of a term is determined by the *actual nature* of the stuff in the samples we use in explaining the meaning of the term together with the *actual content* of our referential intentions. By construction of the example, some kinds of actions are called "wrong" in Twin English by the Twin Earthlings that are not called "wrong" in English by the Earthlings. For example, the Twin Earthlings' use of "wrong" conforms by and large to a deontological theory T_D , and, as we have assumed, this theory implies that lying is wrong. The Earthlings' use of "wrong" conforms largely to a consequentialist theory T_C , and, as we have assumed, this theory does not imply that lying is wrong except in circumstances where lying would have bad consequences. The Twin Earthlings might teach the meaning of the Twin English term "wrong" by, among other things, pointing to a case of lying and saying, "That action and relevantly similar actions are wrong." The Earthlings might teach the meaning of the English term "wrong" by, among other things, pointing to a corresponding instance of lying and saying, "That action is not wrong, for it has good consequences." It follows, Horgan and Timmons would say, that the actual nature of the actions that are called "wrong" in English differs from the actual nature of the actions that are called "wrong" in Twin English. But the Earthlings and Twin Earthlings have relevantly corresponding intentions in using these terms. Both presumably intend something like this: to refer to actions with the relevantly same nature, or that are of the "same moral kind," as the local samples of actions that are called "wrong" in our language.³⁶ If the Earthlings and Twin Earthlings use their terms with intentions that correspond in this way, then since the nature of the actions in the samples on Earth differs from that of the actions in the samples on Twin Earth, it follows in Putnam's theory that the terms refer to different kinds of actions and express different properties. Putnam's theory cannot capture our intuition. This is presumably the argument

35 See Boyd 1988, 210–211.

36 The phrase "same moral kind" is from Sayre-McCord 1997.

Horgan and Timmons would use against a Putnamian version of synthetic naturalism.

The most important error in this reasoning is that it does not take proper account of the strategies that are available to Putnamian semantics for dealing with mistakes. Speaking of “wrong actions” might be like speaking of “unicorn horns.” Calling a lying action “wrong” might be like calling some gin “water.” Or denying that a lying action is “wrong” might be like denying that female dogs give “milk.”

The second error is that the reasoning does not take into account the fact that, for Putnam, some similarities among the items that are included in the extension of a term by speakers might be irrelevant to determining its extension. Putnam says that “important” similarities are what matter and, he says, the notion of importance is “interest relative.”³⁷ To see the point, recall that all so-called unicorn horns are narwhale tusks. This is not an important similarity given our referential intentions and our interests. For a better illustration of the point, suppose that humans had been the only kind of mammal. Then there would have been only one kind of milk and all milk would have had (basically) the same chemical composition. Given our interests in using the term “milk,” however, I propose that the term still would have expressed the functional and genetic property I mentioned before, the property of being a liquid that is produced by the mammary glands of a female mammal that has recently given birth in order to nourish its young. This proposal could be tested by means of a suitable Milk Twin Earth thought experiment.³⁸ If I am correct, then a Putnamian semantics for “milk” would take into account both the presumed referential intention to refer to whatever has the relevantly same nature as the local samples and the similarities that are relevant to our interests, which, again, are functional and genetic similarities. The lesson is that, in applying the semantics, we need to fix on respects of similarity that are semantically relevant, given speakers’ intentions and interests.

There are therefore two key matters that need to be resolved in order to see how Putnamian semantics could be applied to moral terms. First is the content of speakers’ referential intentions in using corresponding English and Twin English moral and twin-moral terms. Second is the kinds of

37 Putnam 1975b, 239, 241.

38 Imagine that the human secretion called “milk” on Twin Earth is chemically different from the human secretion called “milk” on Earth. It is basically XYZ, while the stuff on Earth is basically water. Even so, if the secretion in question on Twin Earth is a liquid with the property of being produced by the mammary glands of a woman who has recently given birth in order to nourish her infant, then my intuition is that it is indeed milk.

similarities that would be semantically relevant, given the interests with which speakers use these terms. We will have to infer these matters from the details given in Horgan and Timmons's construction of the Twin Earth example together with our semantic intuitions and Putnam's theory. It is important that Horgan and Timmons stipulate that the relevant terms in English and Twin English are terms of *morality*. They are not, for example, terms of etiquette or classificatory terms in some social science. Moreover, by construction of the Twin Earth example, Earthlings and Twin Earthlings have corresponding intentions and interests in using the corresponding moral terms in their languages, and the relevant intentions and interests of the Earthlings in the example are the same as our intentions and interests in the actual world in situations in which we use the term "wrong" or its equivalent. This means we can proceed by investigating what Putnam's semantics would or should say about our own semantic intentions and interests in using moral terms.

The place to begin, I think, is with the question, What is presupposed in 'ostensive definitions' of moral terms? Putnam says that ostensive definition has certain empirical presuppositions and that when these presuppositions are false, "fallback conditions" come into play. For instance, he says that my "ostensive definition" of "water" presupposes that the liquid in the sample I use is the same kind of liquid as most of the stuff that I and most speakers in my linguistic community have been calling "water." If this presupposition is false, Putnam says, then "I do not intend my ostensive definition to be accepted."³⁹ Suppose, then, that an English speaker on Earth is teaching a child about right and wrong beginning with the lesson that lying is wrong. Suppose he says, "Lying is wrong."⁴⁰ This teaching does *not* presuppose that the speaker believes lying to be of the same kind or even the same moral kind as most of the acts that the speaker and most speakers in his linguistic community have been calling "wrong." For he might be a nonconformist who disagrees with most speakers in his community about which kinds of acts are wrong. A nonconformist does not have linguistic or semantic intentions different from moral conformists. His moral terms do not have different meanings. And his belief that most people are mostly mistaken about which acts are wrong would not undermine his view that lying is wrong. So he would

39 Putnam 1975b, 225.

40 He would say something like "Lying is wrong," rather than "Lying is called 'wrong,'" because the latter way of speaking would suggest doubt as to whether lying is actually wrong.

not have the intention to withdraw his teaching if it turned out that lying is not of the same (moral) kind as most of the acts that most speakers in his linguistic community have been calling “wrong.” If he is sincere, then of course he *would* intend to withdraw his teaching if he came to believe that lying is *not* wrong. And if he is sincere, he would be committed to withdrawing his teaching if he came to believe that lying is not of the kind, or does not have the property, that he and most speakers in his linguistic community intend to refer to in using “wrong.” This is the kind we call “wrong action,” and the property is the one we call “wrongness.” As Mackie might say, it is the property of “not-to-be-doneness.”⁴¹ Now, in Putnam’s terms, all of this implies that “ostensive definitions” of moral terms presuppose that there are moral properties or kinds and that speakers in the linguistic community use and intend to use moral terms to refer to them. If we came to believe these presuppositions to be false, we would not engage in moral discourse and moral teaching.

In Putnam’s view, this point about the presuppositions of moral discourse implies that we have corresponding referential intentions. For example, it implies that, in using “wrong” (in a nondeviant case), an English speaker intends to refer to actions that are of the kind, or that have the property, that he and most speakers in his linguistic community *intend* to refer to in using “wrong.” Or, more simply, in using “wrong,” an English speaker intends to refer to actions that are of the kind, *wrong*, or that have the property of *being wrong* or of being *not-to-be-done*. And speakers of Twin English would have to have the *same* intentions. If they did not, the relevant terms in their language would not be moral terms.⁴² If speakers used a term, “twong,” for example, with the intention merely to refer to the same kind of action that most speakers in the linguistic community have in fact been *calling* “twong,” then “twong” would not be a moral term. It might be a term of etiquette with a meaning similar to “impolite” or a descriptive classificatory term in the way that “skipping” is a descriptive term for a kind of action. But it would not be a moral term.

Our conclusion so far is perhaps not very interesting considered by itself. The important point, however, is that the argument puts us in a position to understand moral discourse in a way that leaves open the possibility of widespread mistakes about the extension of the term “wrong” in the linguistic community. For even if speakers on Earth believe that all

41 Mackie 1977, 33, 38–40.

42 My position here is similar to a claim made by Mackie about speakers’ intentions in using moral language (1977, 33).

and only actions prohibited by the consequentialist theory T_C are wrong, and even if all the actions *called* “wrong” on Earth are so prohibited, it does not follow from Putnam’s theory, given the referential intention with which we use the term, that “wrong” refers to the kind, actions that are prohibited by the consequentialist theory T_C , or that wrongness is the property of being so prohibited. The argument I reconstructed for Horgan and Timmons turns on assuming that this does follow from Putnam’s theory. This, then, is the first error in the argument. It is the major error.

Let me now turn to the second matter that I said needs to be investigated, the ‘interests’ with which speakers of English use the term “wrong.” On Putnam’s account, these interests determine the kinds of similarities among actions that would be semantically relevant. By construction of Horgan and Timmons’s thought experiment, as I said before, the English term “wrong” and the Twin English term “wrong” are both moral terms. Also by construction, Twin Earthling judgments about what is “right” and “wrong” play the same role in Twin Earthling reasoning, deliberation, and decision making about what to do that Earthling judgments about what is “right” and “wrong” play in the reasoning, deliberation, and decision making of Earthlings. For example, the Earthlings and Twin Earthlings are normally disposed to do what they take to be right or twin-right, respectively; they normally take considerations about right or twin-right, respectively, to be especially important in deciding what to do, and so on. The Twin Earthlings’ beliefs about twin-right and twin-wrong, like the Earthlings’ beliefs about right and wrong, are normative beliefs. Moreover, “wrong” is the primary negative term of act evaluation in the Twin Earthlings’ moral code just as the Earthling term “wrong” is the primary negative term of act evaluation in the Earthlings’ moral code.⁴³ These facts about the example explain, I think, why we have the intuition that the Earthlings and Twin Earthlings disagree in the lying case about the truth of the proposition that lying is wrong. Moreover, they point us in the direction of the interest speakers have in using the term “wrong” in English, or the corresponding term in Twin English. Their interest is to pick out the kind of action or property of actions, whatever that might be, that is of primary importance morally in deciding which actions to

43 By stipulation, the Twin Earthlings’ use of “wrong” complies by and large with a deontological theory T_D , and since the primary negative term of act evaluation in deontological theories is “wrong,” their term “wrong” is the primary negative term of act evaluation in the Twin Earthlings’ moral code. Similarly, Earthling usage by stipulation conforms largely to a consequentialist theory T_C , and the primary negative term of act evaluation in consequentialist theories is also “wrong.”

avoid – the kind that is especially to be avoided, or the property of being morally “to-be-avoided” or “not-to-be-done.”

I am finally in a position to argue that, in any coherent Moral Twin Earth example of the sort described by Horgan and Timmons, Putnamian semantics implies that the English term “wrong” means the same as the Twin English term “wrong.” I have already argued that Earthlings and Twin Earthlings have relevantly corresponding referential intentions and interests in using these terms. The Earthlings’ referential intention in using the English term “wrong” is to refer to actions that are of the kind or that have the property that the speaker and most speakers in his linguistic community intend to refer to in using “wrong,” and the Twin Earthlings’ referential intention is the same except that it refers to the Twin English linguistic community and the Twin English term “wrong” rather than the English community and English term. All the relevant intentions correspond in the relevant way. Moreover, as we have just seen, the Earthlings’ and Twin Earthlings’ interests in using these terms are the same. Their interest is to pick out the kind of action or property of actions that is morally to be avoided, or the property of being morally “not-to-be-done” or “to-be-avoided,” in order to make morally defensible decisions. This property or kind of actions must be the same on Twin Earth as it is on Earth. For by construction of the example, there is no morally relevant difference between the two worlds. People in the different worlds have different moral theories, of course. But Earthlings and Twin Earthlings are rational to the same degree and have the same abilities, vulnerabilities, needs, and desires. Moreover, causal laws are the same on the two planets. Hence, any coherent moral theory would say if wrongness is instantiated on the one planet it is instantiated in the other, and vice versa, and that the kinds of actions that are wrong on the one planet are wrong on the other, and vice versa. But if the worlds are the same in all morally relevant respects, and if corresponding terms in the languages spoken on the two worlds are used with relevantly the same referential intentions, and if people’s interests are the same in using these terms, then the terms would be assigned the same meaning in Putnamian semantics.

It might be useful here to consider a Twin Earth scenario involving a term that is not normative in order to show that the issues raised for Putnamian semantics by the Moral Twin Earth argument can also arise for nonnormative terms. Suppose, then, that Earthlings are acquainted with human milk, cow milk, and goat milk and call each of these liquids “milk” in English. They are not acquainted with sheep milk. Twin Earthlings are

acquainted with human milk, cow milk, and sheep milk and call each of these liquids “milk” in Twin English. But they are not acquainted with goat milk. Both Earthlings and Twin Earthlings are aware that female dogs that have recently given birth secrete a fluid that is drunk by their young. Earthlings say in English that this is “milk,” while Twin Earthlings say in Twin English that it is not “milk.” Twin Earthlings deny that it is “milk” because they believe that all twin-milk is white and they have observed that the relevant secretion of female dogs is yellow. Earth and Twin Earth are otherwise as similar to each other and as similar to the actual world as is compatible with the differences already noted.⁴⁴

I have the intuition in this example that “milk” means the same on Earth and on Twin Earth and that the Earthlings and Twin Earthlings disagree about the secretions of female dogs. But whether or not this intuition would be widely shared, the important point for my purposes is that Putnam’s theory can explain it. By construction of the example, the Earthlings and Twin Earthlings have relevantly corresponding referential intentions in using their “milk” terms and they have relevantly corresponding interests in using the terms. By construction, moreover, there is no chemical composition shared by all liquids in question on either planet. But there is nevertheless a functional and genetic property that is shared by all liquids on Earth that the Earthlings call “milk.” This is the property I mentioned before of being a liquid produced by the mammary glands of a female mammal that has recently given birth in order to nourish its young. This property is also shared by all liquids on Twin Earth that the Twin Earthlings call “milk” and it is the property relevant to the Earthlings’ and Twin Earthlings’ interest in these liquids. Therefore, it is plausible that this is the property expressed by the English term “milk” as well as by the Twin English term “milk.” This means it is plausible that the terms have the same meaning in Putnam’s sense. Since the relevant property is also shared by the liquid produced by female dogs and drunk by their young, the Earthlings are correct that this liquid is milk. The Twin

44 Recall that Horgan and Timmons stipulate that in Moral Twin Earth scenarios there is “significant agreement” between the Earthlings’ moral beliefs and the Twin Earthlings’ twin-moral beliefs. That is, it is true by construction of the Twin Earth scenarios that most actions that the Earthlings believe to be wrong, the Twin Earthlings believe to be twin-wrong, and vice versa. I have described the Milk Twin Earth scenario such that, similarly, there is “significant agreement” between the Earthlings’ milk beliefs and the Twin Earthlings’ twin-milk beliefs. It is true by construction of the Milk Twin Earth scenario that most liquids that the Earthlings believe to be milk the Twin Earthlings believe to be twin-milk, and vice versa.

Earthlings are incorrect. Of course, the Twin Earthlings believe that the relevant secretion of female dogs is not milk because it is yellow, but this fact does not rule out my claim that the Twin English term expresses the functional and genetic property in question. A semantics has to allow for people to have mistaken beliefs about the extensions of their terms. This is precisely the point that needs to be stressed here. The Twin Earthlings are mistaken about what is common to all milk.⁴⁵

In Horgan and Timmons's Moral Twin Earth example, the Earthlings accept a consequentialist moral theory, while the Twin Earthlings accept a deontological theory. According to my argument, they therefore disagree about which actions are wrong. In the lying case, the Earthlings say that the lying in question is not "wrong" while the Twin Earthlings say that it is "wrong." According to my argument, either the Earthlings or the Twin Earthlings are mistaken.⁴⁶ But I have not said which is mistaken, and it is compatible with my argument that both are mistaken. For all I have shown, it is possible that there is no property of wrongness and that "wrong" has an empty extension.

The argument I have given is enough, I think, to refute the Moral Twin Earth argument, but it is not enough to provide a complete Putnamian account of the semantics of "wrong." It also is not enough to show that synthetic moral naturalism is defensible. To defend synthetic moral naturalism, one would have to show both that there is such a property as wrongness and that it is a natural property. This is a task for metaethical theory, not a task simply for a general semantics. Putnamian semantics tells us that the referential intentions with which we use the terms "water" and "milk," and our interests in using the terms, together with the relevant natures of the relevant liquids, determine the meaning of "water" and

45 At this point it might be useful to recall a discussion by Saul Kripke of an imaginary case in which it turns out that our belief that all gold is a characteristic yellow color is false. It turns out that all the stuff we have called "gold" is blue when it is seen in normal lighting conditions. Kripke says this does not mean that there is no gold. We use "gold" as a term "for a certain *kind* of thing." He adds, "The kind of thing is *thought* to have certain identifying marks." In my example, the Twin Earthlings think that all milk has the mark of being white, but they are wrong just as, in Kripke's case, people are wrong to think that gold has the mark of being yellow. See Kripke 1980, 118. I owe this reference to Janice Dowell.

46 Horgan and Timmons suggest that "perhaps" the difference between the Twin Earthlings' use of "wrong" and the Earthlings' use of "wrong" is due to the fact that Twin Earthlings tend to experience guilt more readily and to experience sympathy less readily and intensely than do Earthlings. This difference in temperament, together with associated subtle cultural differences, could explain why the Twin Earthlings disagree with the Earthlings when they do. It could also explain why one of the groups makes moral mistakes that are not made by the other group. Horgan and Timmons 1992a, 245; 1992b, 165.

“milk.” But it is chemistry and biology that tell us the natures of the relevant liquids. Similarly, Putnamian semantics by itself does not tell us whether there is a property of wrongness or which property it is. A normative moral theory would purport to tell us which kinds of actions have the property of being wrong, but taken by itself it would not tell us whether there is such a property nor would it tell us whether, if it exists, it is a natural property. We need metaethical theory to tell us these things.⁴⁷

The Moral Twin Earth argument is important for two main reasons. First, it forces moral theory to look closely at the semantics of moral terms. And second, it shows that an adequate semantics for moral discourse must leave open the possibility of widespread mistakes about the extension of moral terms in the linguistic community. I have argued that a Putnamian semantics can be developed in a way that leaves open this possibility. Of course, there are many outstanding problems that would need to be solved in order fully to defend a Putnamian semantics. There are also problems that would need to be solved in order adequately to defend synthetic moral naturalism. As I said, I think that naturalism has difficulty accounting for the normativity of moral terms.⁴⁸ My concern has been to detail aspects of the Putnamian approach to semantics that enable synthetic moral naturalism to escape the challenge of Moral Twin Earth. My goal has been to show that the Moral Twin Earth argument does not succeed in undermining either the doctrine that moral terms have “synthetic naturalistic definitions” or the doctrine of moral naturalism that moral properties are natural properties. I think that naturalists can begin again to issue visas for travel to Moral Twin Earth.

REFERENCES

- Boyd, Richard. 1988. “How to Be a Moral Realist.” In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, N.Y.: Cornell University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- Gampel, Eric H. 1997. “Ethics, Reference, and Natural Kinds.” *Philosophical Papers* 26: 147–163.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.

47 I have proposed such a theory in Copp 1995.

48 For an attempt to account for the normativity of moral terms within a naturalist moral theory, see Copp 1995. See also chapter 8 in this book.

- Horgan, Terence, and Mark Timmons. 1990–1991. “New Wave Moral Realism Meets Moral Twin Earth.” *Journal of Philosophical Research* 16: 447–465.
- . 1992a. “Troubles on Moral Twin Earth: Moral Queerness Revived.” *Synthèse* 92: 221–260.
- . 1992b. “Troubles for New Wave Moral Semantics: The ‘Open Question Argument’ Revived.” *Philosophical Papers* 21: 153–175.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt., Cambridge: Cambridge University Press, 1993.
- Putnam, Hilary. 1975a. “Language and Reality.” In Putnam 1975c: 272–290.
- . 1975b. “The Meaning of ‘Meaning.’” In Putnam 1975c: 215–271.
- . 1975c. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Sayre-McCord, Geoffrey. 1997. “‘Good’ on Twin Earth.” *Philosophical Issues* 8: 267–292.
- Sturgeon, Nicholas L. 1984. “Moral Explanations.” In *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, ed. David Copp and David Zimmerman, 49–78. Totowa, N.J.: Rowman and Allanheld.

Referring to Moral Properties: Moral Twin Earth, Again

Let “moral naturalism” be the view that (1) the central semantic function of moral terms such as “right” and “wrong” is to ascribe moral properties, properties such as rightness and wrongness, and that (2) each of these properties is a ‘natural’ property. Let “synthetic moral naturalism” be the view, in addition, that (3) each moral property could, at least in principle, be ascribed by a predicate couched in nonmoral naturalistic vocabulary, where (4) the corresponding identity claim is synthetic. That is, where “M” is a moral predicate that ascribes a moral property Mness, and “N” is a distinct (perhaps very complex) predicate couched in non-moral vocabulary that ascribes a natural property Nness, if a version of synthetic naturalism implies that Mness is identical to Nness, then it also implies that this is not a conceptual truth.

A fully developed version of synthetic moral naturalism would include a semantics for moral predicates. It would explain what it is that determines which specific natural property is ascribed by a given moral predicate. To avoid being objectionably ad hoc, moreover, synthetic moral naturalism must take on board a general semantic theory of the relation between predicates and properties and apply this theory to the case of moral predicates. Taken together with relevant nonsemantic facts, the resulting moral semantics should determine (nontrivially) which natural property it is

I would like to thank Terry Horgan, Michael Ridge, Mark Timmons, and Jon Tresan for very useful comments on an earlier draft of this chapter. I am especially grateful to Tresan for many helpful discussions of the ideas in this chapter over the past few years. I received important help as well from the members of Tresan’s graduate seminar in the fall of 2005, including especially Ana Maria Andrei, Murat Arici, and Emil Badici, whom I would like to thank for their suggestions.

that a given moral predicate is used to ascribe in standard literal assertoric uses.¹

Terence Horgan and Mark Timmons have argued, however, that this strategy cannot work. Suppose, for a *reductio*, that the moral predicate “right” is a typical property-ascribing predicate. Now imagine that there is a planet that is the twin of Earth. On Twin Earth, the Twin Earthlings speak Twin English and use a predicate “t-right” that corresponds to the Earthling term “right.” Now suppose that the Earthlings affirm the English sentence “It is always right to maximize the general welfare,” while the Twin Earthlings affirm the Twin English sentence “It not the case that it is always t-right to maximize the general welfare.” We would intuitively say that the Earthlings and Twin Earthlings thereby express a disagreement about what is right. The trouble, Horgan and Timmons argue, is that, for any standard semantic theory, it is possible to devise a scenario of this kind in which the theory would say that the Twin Earthling predicate “t-right” ascribes a property different from the property ascribed by the corresponding Earthling term “right” and that the Earthlings and Twin Earthlings therefore do not genuinely disagree. Given the implausibility of this result, Horgan and Timmons conclude that “right” is not a typical property-ascribing predicate.²

In chapter 6, I claim that this argument, the Moral Twin Earth argument, is not successful in undermining synthetic moral naturalism. Horgan and Timmons complain, however, that I did nothing more in this chapter than to provide “a wish list of what the moral naturalist needs to accomplish.”³ They do not respond point by point to my objections. Instead they shift their ground. They now describe their position as posing a dilemma for any version of synthetic naturalism: Either the theory fails “to fix *determinate* reference-relations between moral terms and certain natural properties,” or the theory is committed to an “objectionable relativism,” in that there are possible circumstances in which two groups “appear to utter contradictory moral statements about some object of moral evaluation,” yet the theory implies that they do not have a “genuine disagreement.”⁴ Horgan and Timmons used to claim that any version

1 It is trivial, for instance, that “right” refers to rightness. An antirealist expressivist who accepts a deflationist semantics could say such a thing. For a discussion of antirealist-expressivism and deflationism see chapter 5.

2 This is my interpretation of their argument. See Horgan and Timmons 1992a, 1992b, and 1990–1991. Also Horgan and Timmons 2000, 141.

3 Horgan and Timmons 2000, 149.

4 Ibid., 140, 143.

of synthetic naturalism will fall victim to the Twin Earth argument.⁵ Now they claim that any version of synthetic naturalism that provides a determinate semantics for moral terms will fall victim to the Twin Earth argument. In effect, their claim is that it is not possible to fill in the details of a semantics for moral terms without running afoul of their dilemma.

Although I do not think that Horgan and Timmons's argument is successful, it is important because it challenges moral naturalists to look more closely at the semantics of moral terms.⁶ In chapter 6, I admit that I do not provide a fully developed account of the semantics of moral terms. I propose that naturalists might adopt a semantics for moral terms that is suggested by Hilary Putnam's views about the semantics of natural kind terms,⁷ but I leave important details to another occasion. My goal here is to fill in some of the details, and to argue that a Putnamian semantics for moral terms can avoid Horgan and Timmons's dilemma. I do not mean to claim that the Putnamian semantics is correct. I am not certain that it is the best approach. Nevertheless, I shall assume for the sake of argument that a Putnamian theory can be designed that gives a plausible semantics for the kinds of nonmoral predicates Putnam had in mind. I shall then suggest a way to extend the Putnamian semantics to deal with moral terms.

According to what I am calling the Putnamian semantics, there is a family of predicates such that, for each of these predicates, the identity of the property it ascribes in standard literal assertoric uses is determined jointly by, on the one hand, speakers' 'semantic intentions' in using the predicate and, on the other hand, the actual nature of the stuff that speakers use in explaining the use of the predicate.⁸ For example, Putnam thinks that we intend to use "water" to refer to the stuff in the rivers and lakes around us and to any quantity of liquid that has the same basic physical nature as this stuff. Given that the stuff in the rivers and lakes around us is H₂O, Putnam's semantics implies that (to be) water is (to be) H₂O. But Putnam also wants to apply his basic approach to at least some functional terms and artifact terms.⁹ To deal with such terms his theory needs to take into account all relevant aspects of the content of our semantic intentions.

5 Horgan and Timmons 1992b, 167, and 1992a, 248.

6 I do not mean to say that philosophers have not previously paid serious attention to this. For two examples, see Boyd 1988 and Jackson 1998.

7 Putnam 1975b, 229–242, and Putnam 1975a.

8 Putnam 1975b, 224–225.

9 He says similar accounts could be given for "many other kinds of words," including "the great majority of all nouns, and . . . other parts of speech as well" (1975b, 242).

In general, we intend to use a term *t* to refer to whatever has the relevantly same nature as samples used in explaining the use of *t*. However, as Putnam points out, what counts as “relevant” depends on our interests.¹⁰ Although “structurally important properties” or “important physical properties” are normally decisive, in some cases other properties are decisive.

A fully articulated Putnamian semantics would have to explain in general which of our intentions in using a term qualify as *semantic* intentions and how the exact content of these intentions is to be determined. Now, it is common to think of the meaning of a term as determined by the semantic *conventions* governing the use of the term, where the existence of a convention in a group is a matter of the existence of overlapping and interlocking intentions and dispositions of a relevant kind.¹¹ The semantic intentions that Putnam has in mind are presumably intentions that suitably mirror the semantic conventions governing correct and literal use of the term in question. For example, suppose that the semantic convention governing the use of “milk” ensures that it is to be used to refer to the liquid secreted by cows from their udders and to the liquid secreted by female dogs from their teats and to any quantity of liquid that has the same kind of origin and biological function as this kind of stuff. Some people may use the term with deviant idiosyncratic intentions. And some people may use the term without a substantive understanding of the background convention governing its use, intending, in using the term, merely to use it in the way it is standardly used, to ascribe whatever property it is standardly used to ascribe. But people who have a substantive understanding of the relevant convention would presumably intend to use the term to refer to the liquid secreted by cows from their udders and to the liquid secreted by female dogs from their teats and to any quantity of liquid that has the same kind of origin and biological function as this kind of stuff. These people would use the term with the relevant semantic intention. At least this is how I propose to understand the Putnamian view. So, I am proposing, if *t* is governed by a semantic convention according to which *t* is to be used to *A*, then the relevant *semantic intention* for the use of *t* is the intention to use *t* to *A*. Speakers whose intentions in using *t* are semantic intentions in the sense relevant to Putnam’s account are those who use *t* with the intention to *A*.

A fully articulated Putnamian theory would have to explain how the semantic intentions with which we use a predicate, together with the

10 Putnam 1975b, 239, 241.

11 See Lewis 1969.

actual nature of the things we speak about in using the predicate, determine which property it ascribes. For example, given the actual nature of the liquids that are secreted by cows from their udders, by female dogs from their teats, and so on, and given the content of the semantic intention with which we use the term “milk,” the Putnamian view presumably would say that the predicate “milk” ascribes to a liquid the complex functional and genetic property of having been produced by a mammary gland of a female mammal that has recently given birth and of having the biological function of nourishing its offspring.¹² To be sure, the theory is not formulated in a sufficiently precise way to yield determinate results in every case. I think, however, that it is as well developed as any competing semantic view.

As we saw, Horgan and Timmons challenge naturalists to produce a theory that fixes “determinate reference-relations between moral terms and certain natural properties.” I think, however, that a certain indeterminacy is to be expected in semantic theory. First, I think it is plausible that there are uncountably many properties. Moreover, second, we have finite and fallible minds. In addition, third, ostention and description are blunt means at best of singling out a specific property. Given all of this, it seems likely that many (if not most) of the predicates we use to ascribe properties do not manage to refer to a single specific property from among all the properties that exist. As Horgan and Timmons say, there are too many properties that are eligible to serve as referents for our predicates.¹³ If this is a problem, it is a problem for any theory of the semantics of moral predicates¹⁴ and, moreover, for semantics much more generally.

12 *American Heritage Dictionary of the English Language*, 3rd ed. (Boston: Houghton Mifflin, 1992), s.v. “milk.” On the Putnamian account, if the people on Earth use “milk” with the semantic intention suggested in the text, and if the people on Twin Earth use “t-milk” with exactly the same semantic intention, and if the actual nature of the liquids in question, including their biological functions, is exactly the same, and if Earth and Twin Earth are the same in every other relevant respect, then “milk” and “t-milk” express the same property and have the same meaning. Of course the Earthlings and Twin Earthlings might have different theories about the biological function of the liquids in question, and as a result, the Earthlings might think that the liquid secreted from the udders of female goats is “milk,” while Twin Earthlings think that this liquid is not “t-milk.” In this case, the Putnamian account would say that the Earthlings and Twin Earthlings have a genuine disagreement and that one of the groups must be mistaken. My idea, in writing chapter 6, was that a well-designed Putnamian semantics for moral predicates should work in Twin Earth cases in a similar way.

13 Horgan and Timmons 2000, 149.

14 One might think it is not a problem for a noncognitivist moral semantics that denies that moral predicates are standardly used to ascribe properties. Yet such a semantics must

It suggests that property-ascribing predicates are all vague in a certain way.

A variety of ordinary predicates could be used to illustrate the point, including “milk,” “beach,” “planet,” “red,” “inflationary,” “efficient,” and “storm.” For a controversial example, consider the term “water.” Suppose that Putnam is correct that we use “water” with the semantic intention of referring to any quantity of liquid that has the same basic physical nature as the stuff in the rivers and lakes. It is well known, however, that the stuff in our rivers and lakes is not pure H_2O . A small proportion of ordinary water is heavy water, or D_2O , and a small proportion is semi-heavy water, or HDO . All naturally occurring water also contains various chemical impurities and it often contains biological impurities as well. Water found in different rivers and lakes varies in the proportions and kinds of impurities it contains. Hence, given the actual physical nature of the samples of water that are available in our environment, even if the content of our semantic intentions in using “water” is what Putnam seems to suppose, it is not clear that the Putnamian theory implies that “water” ascribes the property of being H_2O . There are too many other properties that are eligible referents for the term. For example, there is the property of being a liquid with a concentration of H_2O above $x\%$, where all our samples have a concentration above $x\%$, and there is the property of either being liquid H_2O with the impurities in Lake Erie or being H_2O with the impurities in Gainesville tap water, and so on, for all our samples. And I think it is clear that there are many other properties that are also eligible referents for the term “water.”¹⁵

Since issues about indeterminacy and vagueness arise very generally, they are not a special problem for moral predicates. For this reason, I

say something about the semantics of these predicates. It might say that they are used to express certain appropriate conative attitudes. The problem is that there are too many types of attitudes. Suppose the theory says that “wrong” is used to express moral disapproval. Precisely which attitude counts as moral approval? This issue corresponds to the issue in cognitivist semantics of specifying precisely which property counts as wrongness. Of course one can always dodge the issue by saying that wrongness is the property ascribed by “wrong” or that moral disapproval is the attitude expressed by standard uses of “wrong.”

- 15 This means it is an idealization to suppose that “water” is a natural kind term. The trouble is that the basic physical nature of liquids that we classify as “water” varies from sample to sample. Michael Ridge pointed out in correspondence that we would not count all the liquids that we classify as water as “pure” water. But if we do not beg the question by assuming water is H_2O , it is an open question whether pure water is pure H_2O . If “water” refers to H_2O , then, presumably, “pure water” refers to pure H_2O . But if not, then not. Intuitively, I think, clear, odorless, potable water qualifies as pure water even though it contains some HDO and even if it contains some salts.

think we will have done well enough if we manage to sketch a semantics for moral predicates that does about as well as the Putnamian semantics for “water” and “milk” and similar terms in avoiding or handling worries about indeterminacy. If we can do this, then we can take ourselves to have avoided the indeterminacy horn of Horgan and Timmons’s dilemma. To do this, however, we need to develop a proposal about the content of the semantic intentions with which speakers use the moral predicates.

For simplicity, let me restrict attention to the term “right” and the corresponding Twin Earthling term “t-right.” I shall assume that these are moral terms, and I shall treat them as equivalent to “morally required” and “t-morally required,” respectively. Now, it turns out that if we make certain (implausible) assumptions about the semantic intentions with which “right” and “t-right” are used, the Putnamian semantics can easily be shown to fall victim to the relativism horn of Horgan and Timmons’s dilemma. For it is possible to devise a Twin Earth thought experiment that yields counter-intuitive relativistic implications. The thought experiment does not defeat the Putnamian semantics by itself, however. What it undermines is the conjunction of the semantics with our assumption about semantic intentions.

Suppose, for example, that the Earthlings and Twin Earthlings intend, in using “right” and “t-right,” respectively, to ascribe the property in virtue of having which an action would be required by their community’s moral theory. That is, the semantic intention in each case could be reported by saying it is the intention of “ascribing the property in virtue of having which an action would be required by this group’s moral theory.” Despite the indexical element in the content of the intentions, speakers in the two groups intuitively have the same semantic intention. But we can assume that the Earthlings and Twin Earthlings accept different moral theories. If we assume that the Earthlings’ theory is consequentialist, an act would be required by their theory in virtue of (say) maximizing the general happiness, and so, on the Putnamian approach, it would turn out that the Earthling term “right” ascribes the property of maximizing the general happiness. And if we assume that the Twin Earthlings’ theory is deontological, then, by similar reasoning, on the Putnamian approach, it would turn out that the Twin Earthling term “t-right” ascribes the property of being required by the relevant deontological rules. On this showing, the Putnamian theory would say that Earthlings and Twin Earthlings do not have a genuine disagreement when the Earthlings say it is always “right” to maximize the general happiness and the Twin Earthlings deny that this is always “t-right.” For given the

theory and given our assumption about semantic intentions, the proposition affirmed by the Earthlings is not the proposition denied by the Twin Earthlings.

The argument depends, however, on an implausible assumption about our semantic intentions. When we say that some action is “right,” we do not intend to ascribe to it the property in virtue of having which an action would be required by our community’s moral theory. We recognize that our community’s moral theory may be mistaken. It is coherent to accept a moral theory that one knows would be widely rejected in one’s community. For these reasons and others, the proposed view of our semantic intentions in using “right” is highly implausible.

My own society-centered moral theory suggests a different account of the semantic intentions with which speakers use moral terms.¹⁶ Obviously, I cannot here attempt to defend the theory, nor even to explain it in much detail. The theory provides an account of the content and truth-conditions of moral propositions. It identifies the property of rightness – that is, the property of being the right action in a context C (in relation to the relevant society S) – with the property of being required by the code of rules, whatever it is, the currency of which in S actually would best contribute to S’s ability to meet its needs – its needs, *inter alia*, for social stability, for peaceful cooperative interaction among its members, and for its members to be able to contribute to the overall flourishing of the society. Call this code the “best code” for the society in question.¹⁷ On this account, the facts as to which actions are right depend on which code of rules actually is the best code for the relevant society. But since different societies have the same basic needs, the best code for one society will be very similar to, if not the same as, the best code for another society, as long as their circumstances are basically the same in all relevant respects.

Which is the relevant society? On my account, this is settled by the pragmatics of moral conversation and thought. In the default case, the society at issue in a given context is the smallest society that embraces

16 See Copp 1995; also Copp 1996, 1997. See also the introduction to this book.

17 That is, where S is the relevant society, and where M is actually the “best moral code” for S, the predicate “right” ascribes, roughly, the property of being called-for-by-M. Of course, this is not a conceptual truth. An alternative, also within the spirit of the view, would be to identify the property ascribed by “right” in such a context with the property of being called-for-by-the-best-moral-code-for-S. In chapter 4, I argue that the first construal is intuitively the more plausible. I am ignoring various complexities, including the possibility that several codes are tied as “best” for a given society. See note 22, below.

the person making or entertaining the claim as well as all the people who are party to the conversation and all the people referred to or quantified over or whom the speakers or thinkers have in mind in the context.¹⁸ So, for example, if an American and a Frenchman are debating whether capital punishment is morally acceptable, the relevant society is, presumably, the very large society that embraces the populations of all Western democracies. If two Americans are wondering whether capital punishment is acceptable, it is likely that the relevant society is simply the society that embraces the population of the United States. If a philosopher is considering whether to accept consequentialism, the relevant society is likely to be the notional society of all rational persons, since it is likely that the philosopher views morality as being of universal scope, prescribing duties incumbent on all rational persons. Of course the default does not always obtain. If you are wondering whether something would be acceptable in a society that is in radically different circumstances from your own society, the relevant society is the other society, not your own society, and not a larger society that embraces both societies.

We are now in a position to combine the Putnamian semantic view with the society-centered theory to arrive at an account of the semantic intentions of speakers in using moral terms. The basic idea would be that “right” is used with the semantic intention of ascribing to an action or a kind of action the property of being required by the code of rules, whatever it is, the currency of which in the society in question would best contribute to the society’s ability to meet its needs. Or, more colloquially, it is the intention of ascribing the property of being required by the code that would work best for “our” society – the code, the currency of which among us would enable us – the society comprising all of us – to cope as well as can be with our common problems. The idea then is that the facts about the nature of the relevant society and the circumstances it faces determine which code is actually the best code for the society and so determine which property is ascribed by “right.”

The question I now want to address is whether a Twin Earth thought experiment can be devised in which the account I am proposing – the account that combines a Putnamian semantics with the society-centered theory to produce an account of the semantic intentions with which we use moral terms – leads to counter-intuitive relativistic implications.

18 Copp 1995, 221. See Copp 1996. For an account of the relevant concept of a society, see Copp 1995, ch. 7, as well as the introduction to this book.

In Horgan and Timmons's original Twin Earth thought experiment, the only difference between Earth and Twin Earth is supposed to be that, first, the Earthlings accept a consequentialist moral theory, while the Twin Earthlings accept a deontological moral theory, and, second, the Earthlings and Twin Earthlings differ psychologically in whatever way is necessary to explain the fact that they accept these different theories. Because they accept the different theories, the Earthlings say that certain things are "right" that the Twin Earthlings say are not "t-right," and vice versa. The Earthlings say it is always "right" to maximize the general happiness, for example, while the Twin Earthlings deny that this always "t-right." Intuitively, they have a genuine disagreement. The question is, What is implied by our Putnamian account? The answer is that it agrees that the Earthlings and Twin Earthlings have a genuine disagreement. For if we imagine them to be talking about the same circumstances, so that the question of agreement or disagreement can arise between them, then the relevant society in the context is the same. If the relevant society is the same – both when the Earthlings say it is always "right" to maximize the general happiness and when the Twin Earthlings deny that this always "t-right" – then since the semantic intentions of the Earthlings and Twin Earthlings are the same, the theory says the Earthlings affirm the same proposition that the Twin Earthlings deny.

Can we change the scenario in a such way that the Putnamian account will give rise to counter-intuitive relativistic implications? I will consider two strategies. The first is to describe a context in which the society that is relevant to what the Earthlings are saying is different from the society that is relevant to what the Twin Earthlings are saying. The second is to describe a scenario in which the semantic intentions of Earthlings in using "right" are different from the semantic intentions of Twin-Earthlings in using "t-right."

To describe a context in which the relevant societies are not the same, we need only imagine that the Earthlings say it is always "right" to maximize the general happiness, having only fellow Earthlings in mind, and that, meanwhile, on Twin Earth, the Twin Earthlings deny that it is always "t-right" to maximize the general happiness, having only fellow Twin Earthlings in mind. In this case, if the circumstances on Twin Earth differ from the circumstances on Earth in such a way that different moral codes qualify as "best" for the different societies, then the theory says that the Earthlings affirm a proposition different from the proposition the Twin Earthlings deny, so they do not genuinely disagree in what they say. If,

intuitively, they *do* disagree in what they say, we seem here to have a counter-intuitive implication of the account.

Yet I think it is not clear what our intuitions should lead us to say about the scenario. Intuitively, we hear the Earthlings as asserting that it is always “right” (in any context and for any agent) to maximize the general happiness. This is something we hear the Twin Earthlings as denying, so we hear the Earthlings and Twin Earthlings as disagreeing. But if the Earthlings actually said what we intuitively hear them to be saying, or if they actually had in mind every possible agent, then the Putnamian account would agree with our intuitions and imply that the Earthlings and Twin Earthlings disagree. In the scenario we have imagined, however, the Earthlings intend simply to be speaking of Earthlings and it is left open whether the Earthlings have ever considered whether it would be “right” for non-Earthlings to maximize the general happiness. Since the Earthlings intend simply to be speaking of Earthlings when they say that it is always “right” to maximize the general happiness, it seems plausible that they are expressing a proposition about what would be right for Earthlings to do. And since the Twin Earthlings intend simply to be speaking of Twin Earthlings when they deny that it is always “t-right” to maximize the general happiness, it seems plausible that they are expressing a proposition about what it would be right for Twin Earthlings to do. Hence, it seems intuitively plausible to me that the Earthlings are not affirming the proposition the Twin Earthlings are denying. The scenario we have devised is not one in which the account has clearly counter-intuitive implications.

The second strategy is to describe a scenario in which the semantic intentions of Earthlings in using “right” differ from the semantic intentions of Twin Earthlings in using “t-right.” But it would beg the question against the Putnamian account to describe a scenario this way if the scenario is one in which we have the intuition that the Earthling term “right” means the same as the Twin Earthling term “t-right.” For given this intuition, the Putnamian account would take the Earthlings and Twin Earthlings to have the same semantic intentions in using their terms.

Of course we can *stipulate* that the Twin Earthlings use “t-right” with semantic intentions that differ from the semantic intentions with which the Earthlings use “right.” But this is not a way to show that the Putnamian account has counter-intuitive implications. For if we keep clearly in mind the stipulation that the Twin Earthlings use “t-right” with semantic intentions that are relevantly different from the semantic intentions with which the Earthlings use “right,” our intuitions cannot help but be affected if we

are thinking clearly. If our intuitions are properly informed, then given our stipulation, we ought to find it plausible that the Earthlings and Twin Earthlings are not expressing genuine disagreement when, for instance, the Earthlings say it is always “right” to maximize the general happiness and the Twin Earthlings deny that this is always “t-right.” This is because, by stipulation, their *semantic intentions* are different. Given this stipulation, it makes no sense to suppose that what the Earthlings affirm is exactly what the Twin Earthlings deny.

One might object that we can stipulate that there are differences between the intentions with which the Twin Earthlings use “t-right” and the intentions with which the Earthlings use “right” without specifying that the intentions in question are *semantic intentions*. The trouble with this response, however, is that to make difficulty for the Putnamian account, the imagined difference in intentions will have to be one that commits the account to ascribing relevantly different semantic intentions. This is not, then, a way to avoid my line of reasoning.

There is a better objection. One might point out that discourse and thought on Twin Earth about what is “t-right” plays the same kind of practical role in the life of the Twin Earthlings that discourse and thought about what is “right” plays in the life of the Earthlings. Moreover, the terms are used to express the same attitudes, namely, a kind of approval.¹⁹ This means, among other things, that the Earthlings and Twin Earthlings pay the same kind of attention to their respective beliefs about what is “right” and about what is “t-right” in deciding how to act. Hence, when the Earthlings say that maximizing the general happiness is “right” and the Twin Earthlings deny that this is “t-right,” they thereby express a disagreement about how to act, and because of this it is reasonable for us to take them to express a genuine moral disagreement. Because of this, the theory’s implication that they do not disagree will seem counter-intuitive even if we bear in mind the stipulation that the Earthlings and Twin Earthlings have relevantly different semantic intentions. So goes the objection.²⁰

I agree that we might find our intuitions pulled in both directions. On the one hand, bearing in mind the stipulation about semantic intentions, if we are thinking clearly, we will find the implication of the Putnamian account that the Earthlings and Twin Earthlings are not expressing a genuine disagreement to be plausible. On the other hand, bearing in

19 A moral naturalist can allow that moral terms have an expressive role; see chapter 5.

20 Jon Tresan urged me to discuss this objection. I present similar considerations in chapter 6.

mind the similar practical and expressive roles of “right” and “t-right,” we might find this implication of the account to be counter-intuitive. But the account can accommodate our conflicting intuitions once we take into account what is involved in *translation*.

I discussed this point in detail in chapter 6. The point is that the Twin Earth thought experiment in effect treats English and Twin English as different dialects, which raises the question of how best to translate “t-right” into standard English. Now, the best translation of a term from one language or dialect into another is not in general a term that has precisely the same meaning. In some cases there is no term in the home language or dialect that has precisely the same meaning as the term that is to be translated. This means that even if “right” and “t-right” do not have exactly the same meanings – as they do not, according to the Putnamian account, given our stipulation that the terms are used with different semantic intentions – it still might be the case that “right” is the best translation into English of the Twin English term “t-right.” It is therefore open to the Putnamian account to explain our intuition that the Earthlings and Twin Earthlings disagree about what is “right” on the basis of a fact about translation: The Twin Earthlings deny that maximizing the general happiness is “t-right,” where this is best translated into English as the denial that maximizing the general happiness is “right,” but the Earthlings *affirm* that maximizing the general happiness is right. So they are best translated as disagreeing. This is why we are reasonable to take them to disagree.

A plausible translation of “t-right” into standard English would pay attention to the semantic intentions with which (we imagine) “t-right” is used as well as to the practical and expressive role of the term in the discourse and thought of the Twin Earthlings. Suppose, for example, that we imagine the Twin Earthlings to use “t-right” with the intention of ascribing the property of being in accord with the local conventions. In this case, the best translation of “t-right” into standard English would be “conventional,” and if so, we would not or should not think the Twin Earthlings disagree with the Earthlings when Twin Earthlings deny that it is “t-right” – that is, conventional – to maximize the general happiness and the Earthlings insist that it is right to maximize the general happiness. In the imagined scenario, the Twin Earthlings treat the local conventions as overriding, and they approve of action in accord with the conventions, but, for all that has been said, they might have no idea of a standard of rightness against which the local conventions could be evaluated. For all that has been said, they might be puzzled by what the Earthlings say, assuming they know that the Earthlings realize that maximizing the general happiness is

not conventional. The Twin Earthlings might wonder, “What could the Earthlings mean by saying that maximizing the general happiness is ‘right’ since they know it is not t-right?”

Horgan and Timmons would no doubt like to press the worry about indeterminacy at this point. They might question, for example, whether there is in general a fact of the matter as to which moral code is the “best code” for a given society. It seems to me, however, that even if there is not a code that is strictly speaking best for a society S, there will be a (non-empty) set of codes that are tied as “best” for S – in the sense that (a) each of the codes in the set is such that its currency in S would contribute discernibly better to S’s ability to meet its needs than would the currency of any code outside the set and (b) none of the codes in the set is such that its currency in S would contribute discernibly better or worse to S’s ability to meet its needs than would the currency of any other code in the set.²¹ Call the codes in this set “S’s best codes.” The society-centered theory needs to be amended to allow for the possibility of multiple codes being tied as best for a society S. It should identify the property of being right (in relation to S) with the property, roughly, of being required by all of S’s best codes.²² And the Putnamian account should take the semantic intention with which “right” is used to be the intention, where S is the relevant society, of ascribing the property of being required by all the codes, the currency of which in S actually would contribute as well as any others to S’s ability to meet its needs. More loosely, this is the intention of ascribing the property of being required by the codes that would work best for “our” society – the codes the currency of which among us would enable our society to cope with our common problems. It seems to me that these modifications of the account take care of the worry that there may not be a code that is strictly-speaking best for a society.²³

I have already argued that there is a general problem of semantic indeterminacy in philosophy of language. Any account of the semantics of moral terms faces the problem, whether the account is naturalistic, non-naturalistic, or a form of antirealist-expressivism. On the view that “right” ascribes a nonnatural property, we can ask exactly which nonnatural

21 I write “discernibly better” because measurement in this context cannot be expected to be precise.

22 I propose an amendment to the society-centered theory to deal with ‘ties’ in Copp 1995, 198–199 and Copp 1996, 257–258. See also the introduction to this book. Recall that for these purposes I am treating “right” as equivalent to “morally required.”

23 There is a related issue about moral ‘indeterminacy’ that I discuss in Copp 1995, 231–235, and Copp 1996.

property it ascribes, and how exactly the semantics works such that “right” refers specifically to that property from among the entire inventory of properties. On the antirealist-expressivist view according to which “right” is used to express a conative attitude, we can ask exactly which attitude it expresses, and how exactly the semantics works such that “right” comes to be used to express precisely that attitude, given the variety of attitudes that can be found in the human heart. There are of course uninformative answers, such as that “right” ascribes rightness, but such answers are available to naturalism as well as its competitors. There is not a special problem here for the kind of semantics needed by the moral naturalist.

Even if the society-centered theory can avoid Horgan and Timmons’s dilemma, there are of course a variety of other objections to it. One might object, for example, that it is an open question whether morally right actions are required by the code of rules, whatever it is, the currency of which actually would best contribute to the relevant society’s ability to meet its needs. This is a version of G. E. Moore’s “open question argument,” and I have discussed it elsewhere.²⁴ Or one might object that the relativism in the theory – or its “relationalism”²⁵ – is by itself counter-intuitive, and that we do not need to devise Twin Earth scenarios to see this. I have also addressed this complaint elsewhere.²⁶ Obviously, however, this is not the place to attempt to defend the society-centered theory from all such objections.

It is important for a metaethical theory to provide an account of the semantics of moral terms that fits smoothly into a general semantic theory. Moreover, it is important for synthetic moral naturalism to provide a semantics that determines (as exactly as can reasonably be expected) which natural property is ascribed by given a moral predicate. Horgan and Timmons think that this cannot be done by a theory unless it falls victim to the relativism horn of their dilemma, but I think that their pessimism is unwarranted. In this chapter, I have sketched a Putnamian semantics that takes up a suggestion derived from my own society-centered theory about the content of the semantic intentions with which we use moral terms. I have argued that this semantics avoids both horns of Horgan and Timmons’s dilemma. It avoids indeterminacy as much as can reasonably be

24 For Moore’s argument, see Moore 1903, sec. 13, at pp. 67–68. For my response, see Copp 1995, 230–231; also chapter 5 in this book. See also Smith 2003, 194–202.

25 Peter Railton suggested in discussion that the theory is better understood as “relational” rather than “relativistic,” given how the term “relativism” tends to be understood.

26 Copp 1995, 218–223; Copp 1996, 260–262. See also the introduction to this book.

expected and it distinguishes, in an intuitively plausible way, between cases of genuine moral disagreement and cases in which apparent disagreement is merely verbal.

REFERENCES

- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, N.Y.: Cornell University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In *Moral Knowledge? New Readings in Moral Epistemology*, ed. Walter Sinnott-Armstrong and Mark Timmons, 243–266. New York: Oxford University Press.
- . 1997. "Does Moral Theory Need the Concept of Society?" *Analyse und Kritik* 19: 189–212.
- Horgan, Terence, and Mark Timmons. 1990–1991. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16: 447–465.
- . 1992a. "Troubles on Moral Twin Earth: Moral Queerness Revived." *Synthese* 92: 221–260.
- . 1992b. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived." *Philosophical Papers* 21: 153–175.
- . 2000. "Copping Out on Moral Twin-Earth." *Synthese* 124: 139–152.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Lewis, David. 1969. *Convention*. Cambridge, Mass.: Harvard University Press.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt., Cambridge: Cambridge University Press, 1993.
- Putnam, Hilary. 1975a. "Language and Reality." In Putnam 1975c: 272–290.
- . 1975b. "The Meaning of 'Meaning.'" In Putnam 1975c: 215–271.
- . 1975c. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Smith, Michael. 2003. "Moral Realism." In *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-ethics*, 181–207. Cambridge: Cambridge University Press.

Part Three

Naturalism and Normativity

Moral Naturalism and Three Grades of Normativity

Is moral naturalism ruled out by the fact that morality is normative? I want to consider this question in a systematic way, explaining the central thesis of moral naturalism as I understand it, and then clarifying the idea of normativity. The chief point that I want to make is that the issue raised by this question is much more complex than it might seem to be, mainly because of complexity in what philosophers have in mind when they speak of morality as “normative.” There are several dimensions to the complexity, but the dimension I will stress is that philosophers disagree about what we might call the ‘stringency’ of moral normativity. I will distinguish three ‘grades of normativity.’ I will investigate, for each of these grades, the plausibility of the idea that morality has that grade of normativity, and the viability of naturalistic accounts of it. My conclusion will be tentative, but it will be optimistic for moral naturalism.

Moral *naturalism* holds that in thinking of things as morally right or wrong, good or bad, we ascribe moral properties to these things – properties such as moral rightness and wrongness, goodness and evil. It holds that there *are* such properties, and it adds that these properties are ordinary garden-variety natural properties – properties that have the same basic metaphysical and epistemological status as the properties a tree can

I presented an early version of these ideas in September 2003 at a conference on Naturalism and Normativity at the Ethik-Zentrum of the University of Zurich; in October 2003 at the department of philosophy at Ohio University; and in March 2004 at the St Louis Philosophy of Social Science Roundtable. I benefited greatly from the helpful discussion on these occasions. I am also grateful to Philip Clark, Sarah Buss, and Evan Tiffany for very helpful correspondence by e-mail that helped me to clarify my thinking about authoritative normativity. I am especially grateful to my colleagues in the Florida ethics reading group who discussed this chapter: Peter Barry, John Biro, Marina Oshana, Crystal Thorpe, Jon Tresan, Anton Tupa, and Gene Witmer.

have of being deciduous, and the property a piece of paper can have of being an Australian twenty-dollar bill. I will have more to say about this in what follows. The question is whether moral naturalism can accommodate the *normativity* of morality. This is basically the question whether it can account for the fact that morality is, in a characteristic way, action-guiding and choice-guiding. Moral thought and discourse concern how we are to act, what we are to choose, and how we are to live; they involve us in evaluating, prescribing, and recommending. I use the term “normative” to speak of this phenomenon.¹

Unfortunately, the precise nature of the phenomenon, and how we might account for it, are both unclear and contested. Anti-naturalists tend to view naturalist proposals as inadequate, mainly, I think, because they view them as too weak to account for the importance of morality, especially its importance to rational decision-making and action. I therefore attempt in this chapter to distinguish different proposals about what the normativity of morality might consist in. These are the three ‘grades of normativity’ that I mentioned, but one might think that the grades can be subdivided, or one might think there are additional grades. There is also room to disagree about how best to understand each of the grades that I propose. Most important, there is no agreement that morality has all three grades of normativity. Hence, to show that moral naturalism cannot accommodate the fact that morality is normative, one would have to specify which grade of normativity is at issue, to argue that morality does have that grade of normativity, and then to argue that moral naturalism cannot account for this.

I begin the chapter with an attempt to motivate the view that moral naturalism cannot account for normativity. We obviously cannot make genuine progress, however, unless we have a clear grasp of what is meant both by “moral naturalism” and by “normativity,” so this is where I then turn. In the [second section](#) of the chapter, I provide a brief explication of moral naturalism. In the [third section](#), I provide an initial, pre-theoretical explication of the three grades of normativity, which I call *generic*, *motivational*, and *authoritative* normativity. In the rest of the chapter, I discuss the two central questions about each of the grades: Does morality have that grade of normativity? And can naturalism account for it? I think that morality does possess generic normativity, which is the weakest grade of

1 I want to avoid the term “prescriptive” because it has been given a technical meaning in Hare 1952.

normativity. But I argue that it is not plausible that morality has either motivational or authoritative normativity. If I am correct, then even if for some reason there cannot be naturalistic accounts of motivational or authoritative normativity, this would not argue against naturalism. I claim, nevertheless, that it *is* possible to provide naturalistic accounts of all three grades of normativity. In short, my conclusion will be optimistic for moral naturalism. At the very least, I aim to show the difficulties standing in the way of an argument from normativity against naturalism.

1. THE PROBLEM

Why might one think that naturalism cannot account for normativity? Let me ignore, for a moment, the complexity introduced by the idea that there are different grades of normativity.

Since naturalism holds that the moral properties are natural properties, the place to begin is with arguments claiming to show that no natural property can be normative. An anti-naturalist might invoke a version of G. E. Moore's open question argument.² It appears that whatever natural property we might consider, it is an open question whether the fact that something instantiates the property deserves to be given any special significance in our decisions and choices. Normative properties, however, it may seem, are such that there is no question but that things that instantiate them deserve to be given special significance in decision-making. For example, the wrongness of a contemplated action must be taken into account in deciding what to do. It seems to follow that no natural property is normative. Alternatively, an anti-naturalist might appeal to J. L. Mackie's argument from queerness. Mackie argued that a property with "to-be-doneness" built into it would be "queer," and that such a property would be "utterly different from anything else in the universe."³ One might take this to mean that moral properties must be nonnatural, if there are any such properties.

In response, a naturalist might object to the accounts of normativity that are implicit in these arguments. She might claim that there actually *is* an open question whether to pay attention to moral considerations, and so, if moral properties are normative, the key premise of the

2 Moore 1903, sec. 13, at pp. 66–68. Moore intended the open question argument to rule out the thesis that the property of goodness is "complex."

3 Mackie 1977, ch. 1.

Moorean argument must be rejected. Or she might claim that there is nothing queer about normativity, once it is properly understood. She might exhibit a naturalistic analysis of normativity and then argue that normativity so understood is compatible with naturalism. But there are other strategies the naturalist might consider as well, strategies that attempt to account for normativity without providing an analysis, or that attempt to accommodate the normativity of morality without supposing that moral properties are normative.

A naturalist might contend, for example, that normativity is best seen as consisting in a relation between moral *thoughts* and the nature of rational agency, viz., that rational people are motivated appropriately by their moral beliefs. Or she might contend that it is the moral *concepts* that are normative, not the properties they represent. Or she might suggest that normativity is a property of moral *speech acts*, not of moral properties. These strategies may seem unsatisfactory, for it may seem that in making a moral judgment we are responding to something normative that is independent of us, something that is the object of our judgment. It might seem that we cannot adequately account for this without supposing the moral properties themselves to be normative. It is nevertheless important to recognize the availability of these other strategies. Expressivist opponents of moral naturalism adopt strategies of these kinds.⁴ The naturalist should not be in any greater difficulty in taking up these strategies than her opponents. I will return to this issue.

These different strategies for defending naturalism lie along one dimension of the complexity that an anti-naturalist argument must navigate. Along this dimension are different views about what exactly it is that is characterized by normativity.⁵ In what follows, to simplify matters, I will tend to blur distinctions along this dimension by speaking of the normativity of 'morality,' or of 'moral claims,' except when I need to bring the discussion into sharper focus.

Along a second dimension of complexity are different strategies a naturalist could adopt in order to show that normativity is a natural phenomenon. Assume that normativity is a second-order property of the moral properties. There are reductive and nonreductive strategies. One familiar reductive approach analyzes normativity as a complex motivational property. Another identifies normativity with a complex

4 See, e.g., Gibbard 1990.

5 In thinking about this dimension of complexity, I was helped by Jon Tresan.

norm-relative or standard-relative property.⁶ In general terms, a reductive strategy specifies a property, using a term or description that would antecedently be agreed to pick out a natural property – perhaps a complex property the constituents of which are natural properties – and it proposes that the property of being normative is identical to that property.⁷ A nonreductive strategy argues that the property of being normative is a natural property, but attempts to do this without proposing a reductive identity statement.⁸ In order to show that moral naturalism cannot accommodate the fact, if it is a fact, that moral properties are normative, one would have to show that no naturalistic strategy of any of these kinds could be successful.

In what follows, instead of attempting to answer anti-naturalist arguments directly, I will consider each of the three grades of normativity, and ask how a naturalist might attempt to account for the idea that morality has that grade of normativity. If I am correct, it will become clear that the anti-naturalist arguments are unsuccessful.

2. MORAL NATURALISM

As I use the term, a moral naturalist is a moral realist. She agrees with other moral realists – those we call ‘nonnaturalists’ – in accepting a number of doctrines. First, there are moral properties and relations, such as the properties of rightness, wrongness, goodness, virtuousness, and the like.⁹ Second, although moral properties differ from nonmoral properties in that, after all, they are moral properties and the others are not, they have the same basic metaphysical status that any property has – whatever that status is. Third, one of the chief semantic roles of the moral predicates is to express such properties. For instance, in calling an action “right,” one thing we are doing is ascribing rightness to the action. Fourth, a moral assertion expresses a moral belief that represents the world as being a certain way, and what is thereby asserted is true if and only if the expressed

6 My own ‘standard-based’ approach is an example of the latter. See Copp 1995, ch. 1. Michael Smith’s proposal is an example of the former kind of approach. Smith denies that his account is *reductive*, but he also denies that a reductive analysis is needed in order to vindicate naturalism. See Smith 1994, 161–164 and 184–186.

7 For an account of philosophical analysis as consisting in the analysis of complex properties, see King 1998. Other recent accounts of analysis are found in Jackson 1998, Smith 1994, and Gibbard 1990.

8 For a discussion of nonreductive naturalism, see Sturgeon 2006.

9 In what follows, for the most part, I include relations as a kind of ‘property.’

belief is true. This would involve, at least in the case of a basic moral belief, the instantiation of a moral property.¹⁰ And finally, fifth, some moral properties are actually instantiated. For instance, some actions are actually right and some are actually wrong.

The moral naturalist differs from other moral realists in holding that the moral properties are ‘natural’ properties. In brief, then, moral naturalism is the view that there are moral properties and that these properties are natural properties.

For our purposes, the most important question raised by the naturalist’s doctrine is, What is meant in this context by a ‘natural’ property? There is room for disagreement about this, and I do not need to insist on any particular account. In introducing the idea, I suggested that natural properties can be taken to be ordinary ‘garden-variety’ properties with the same basic metaphysical and epistemological status as the property a tree can have of being deciduous or the property of being an Australian twenty-dollar bill. This intuitive understanding of the idea of a natural property might be adequate for my purposes in this chapter.

I cannot go into detail here, but I have proposed elsewhere that we should construe the moral naturalist as holding that moral properties are empirical properties.¹¹ The idea of the empirical is itself problematic, but the proposal at least has the virtue of explaining naturalism in terms of a concept that is not special to moral theory. Any adequate epistemology needs to provide an account of the empirical, so my proposal can piggyback on a concept that is widely used and understood, even if it is contestable. To be more specific about my proposal, the basic idea is that, if a property is natural, then empirical evidence is relevant to the justification of our beliefs about its instantiation. Elaborating further, I would propose that a property is natural if and only if – leaving aside conceptual truths – any proposition about the instantiation of the property that can be known can be known only empirically, or by means of empirical observation and standard modes of inductive inference.¹²

10 A *basic* moral proposition is such that, for some moral property M, it entails that something instantiates M. An example is the proposition that capital punishment is wrong. Among nonbasic moral propositions are propositions such as that nothing is morally wrong and that either abortion is wrong or $2 + 2 = 4$. Call a belief a “basic” moral belief just in case the content of the belief is a basic moral proposition.

11 I develop this idea in more detail in chapter 1. For criticism of this conception of a natural property, and discussion of alternatives, see Sturgeon 2006.

12 Even a naturalist may want to concede that some ethical knowledge is not empirical. For example, it seems plausible that the concept of murder is the concept of a wrongful killing, and if this is so, then we can know *a priori* that murder is wrong. It is no part of ethical

This proposal is in line with Moore's basic ideas about moral naturalism. Moore took naturalism to be the view that the property goodness is a natural property,¹³ but he got into difficulty in attempting to explain the idea of a natural property.¹⁴ My proposal is similar to his first brief characterization of naturalism, in section 25 of *Principia Ethica*, where he wrote that, according to naturalism, "Ethics is an empirical or positive science: its conclusions could all be established by means of empirical observation and induction."¹⁵ It clearly would be controversial to hold that ethics is a science. Instead, what the naturalist might say, following Moore's suggestion, is that ethics is empirical in that any substantive ethical knowledge is empirical, or based in "empirical observation and induction." My proposal is in line with this idea.

Given this construal, moral naturalism is not entailed by metaphysical naturalism, for a metaphysical naturalist could deny that there are any moral properties at all, thereby rejecting moral naturalism. Nor is moral naturalism a kind of physicalism. It is left open whether there are natural properties that are nonphysical.

For the most part, I will simply assume the realist position that there are moral properties. The central issue I want to address is whether the normativity of morality somehow rules out the idea that these properties are natural.

3. THREE GRADES OF NORMATIVITY

The basic idea, as I said, is that morality concerns how to act, what to choose, and how to live. It prescribes, recommends, and evaluates actions and choices. This, however, is a characteristic that morality shares with etiquette, prudence, epistemology, aesthetics, and judgments of individual self-grounded rationality. In judging that it is rude to carry on a loud conversation during a concert, we evaluate such conversations. Apparently, then, there is a kind of *generic* normativity shared by these various kinds

naturalism as such to deny that there are conceptual truths in ethics and that our knowledge of any such truths is not empirical.

13 Moore 1903, sec. 26, p. 93.

14 He later acknowledged that his attempts to explain the idea of a natural property in *Principia* were "hopelessly confused" (Moore 1993, 13). He admitted that one of his proposals was "utterly silly and preposterous." He added that in *Principia* he had not given "any tenable explanation" of what he meant by saying that goodness is not a natural property. See Moore 1968, 581–582.

15 Moore 1903, sec. 25, p. 91.

of thought and discourse. But one might think that moral thought and discourse are normative in a further respect. For one might think that when a rational person accepts a moral claim, she takes it into account in deciding how to act or choose. This is a kind of *motivational* normativity that distinguishes morality from etiquette and aesthetics, since rationality does not ensure that one will take into account considerations of etiquette or aesthetics. But one might think that moral thought and discourse are normative in still a further respect. For one might think that there are reasons to do what is morally required even if we are unaware of them. And one might think that a rational person who understood this would not take there to be a genuine question about whether to be moral, or at least about whether to give moral considerations due weight in her deliberation. Morality has *authoritative* normativity, we might say. When philosophers allude to the ‘normativity’ of moral thought and discourse, they could have any of these ideas in mind.

One might wonder how these so-called grades of normativity are related to the traditional Kantian distinction between hypothetical and categorical imperatives. The answer is that there are different conceptions of the categorical. Each of the grades of normativity could be understood as a proposal about what the categorical nature of moral requirements comes to. On one view, generic normativity of a certain kind is sufficient for a requirement to be categorical. On another view, moral requirements are genuinely categorical only if they have authoritative normativity.¹⁶

Some philosophers might be tempted to explicate normativity by invoking the idea of a reason. One problem for this approach is that claims about what we have reason to do are themselves normative, so such an approach does not go very deep. Another problem is that there seem to be reasons of different kinds. There appear to be moral reasons, reasons of etiquette, epistemic reasons, and so on. Moreover, these different kinds of reasons appear to have different grades of normativity. Reasons of etiquette might be less stringently normative than moral reasons, and moral reasons might be less stringently normative than rationally compelling

16 See Foot 1978. Foot pointed out that, on one account of the categorical, the requirements of etiquette qualify as categorical, but that this account would be viewed as too weak by those who think that morality is distinctive in being a source of categorical requirements. Foot denied that moral considerations “necessarily give reasons for acting to any man,” and in so doing she took herself to be denying that morality is a source of categorical reasons (161).

or strongly authoritative reasons – unless, of course, they are themselves strongly authoritative. Because of this, before we could explicate normativity by invoking the idea of a reason, we would need to ask what kind of reason is at issue and what is its grade of normativity. This is not a way to avoid issues about the grades of normativity.

To a moral realist, it may seem natural to explain the normativity of morality by supposing that moral properties have the second-order property of being normative, and, if there are different grades of normativity, by supposing that there are second-order properties corresponding to the different grades. As I mentioned before, however, a realist does not need to ascribe normativity to the moral properties; there are alternatives. Moreover, at this stage of the discussion, I want to provide an intuitive account of the different grades of normativity without presupposing moral realism. Hence, as I said, I will present the grades of normativity as putative properties of ‘morality,’ or of ‘moral claims.’

(1) *Generic Normativity*

To begin, then, compare the claims that lying and deception are wrong with the claim that lying and deception are widespread.¹⁷ There is a characteristic difference between these claims. The nonmoral claim is simply descriptive of aspects of human behavior, but the moral claim is not merely descriptive. The claim that lying is morally wrong is an evaluation of lying and deception. It is a proscription of lying. Moreover, it is evaluative or proscriptive in virtue of what it says. That is, it has these properties in virtue of its semantics, meaning, or content; it is *essentially* recommendatory or proscriptive. To generalize, it appears that moral claims have a characteristic essential semantic connection to decisions about action or choice, a connection that can be referred to by describing them as ‘normative.’¹⁸ Call this property of moral claims “generic normativity.”¹⁹ It is the weakest grade of normativity.

I say that moral claims are not merely ‘descriptive,’ but to a moral realist, basic moral claims express propositions that attribute moral properties to things, and in this sense they *do* describe those things. To a realist, for instance, we ascribe wrongness to torture in saying or thinking

17 In this and the next two paragraphs I draw on Copp 1999.

18 This is not meant as a definition. It is at best a characterization of generic normativity.

19 The term “generic normativity” was suggested by Gene Witmer, in conversation.

that torture is wrong. In addition, however, such claims evaluate, recommend, or prescribe; they have (at least) generic normativity. The problem for the realist, then – whether or not she is a naturalist – is to provide a philosophical account of the generic normativity of moral claims on the assumption that such claims express propositions and ascribe properties.

Various nonmoral claims are also normative in the generic sense, including claims about what we have prudential reason to do or epistemic reason to believe, and claims about what we are required by etiquette to do. There are, in addition, different kinds of moral claims. An adequate theory of generic normativity must account for these other kinds of generically normative claims as well as moral claims, and do so without obscuring the differences among them.

(2) *Motivational Normativity*

One might think that moral thought and discourse are normative in a deeper way than, say, thought and discourse about the requirements of etiquette. A person who thinks she is required in etiquette to do something cannot be convicted of any kind of irrationality even if she views etiquette as beneath contempt and even if she pays it no attention in deciding how to act. Morality seems to be different from this. It has seemed to many philosophers that if a *rational* person thinks she is *morally* required to do something, she will be motivated to do it. More generally, and perhaps more cautiously, a person who is practically rational takes her moral beliefs appropriately into account in deciding how to act or choose, and if such a person believes she is morally required to do something all things considered, she is motivated to do it.²⁰ We can call this alleged property of moral claims “motivational normativity.”

The thesis that moral claims have motivational normativity is a close relative of a stronger view, a view I will call “motivational internalism” or “judgment internalism,” according to which, practical rationality aside, it is a necessary truth that *anyone* who thinks she is morally required to do something will be motivated to do it.²¹ There are familiar objections to motivational internalism, however. It appears that people who are depressed might lack motivation to do what they think they morally

20 For such a view, see Korsgaard 1986; also Smith 1994, 62.

21 The term “internalism” is used for many different positions. The doctrine in question here is the one that Stephen Darwall has called “judgment internalism” (1983, 54–55).

ought to do. And it appears that people with unusual second-order beliefs about morality might also fail to be motivated to do what they think they ought to do. For example, in Graham Greene's *The Heart of the Matter*, Scobie does something he believes to be forbidden in order to secure his own damnation, believing that he is thereby expressing love of God. Perhaps Scobie still has some motivation not to do the forbidden thing, but it is not clear that this is so, and it is doubtful that it must be so. For a different example, suppose someone thinks that God's commands determine what we are morally required to do, but suppose she is in rebellion against God. She could conceivably lack any motivation to do what she thinks morality requires.

There is an additional issue, which is more important for our purposes. It is not clear that motivational internalism is a plausible account of a kind of *normativity*. Consider, for example, that people are motivated to eat things that they take to be sweet. Of course, this is not a necessary truth, and motivational internalism is supposed to be necessarily true. But we can fix this asymmetry by stipulating that a person counts as believing something is *tweet* if and only if she believes it is sweet and in addition she is motivated to some degree to eat it. It is a necessary truth that a person who thinks that something is *tweet* is motivated to eat it, but I doubt that anyone would think on this basis that the claim that something is *tweet* is *normative*. There is a motivational condition on the sincere use of "tweet" that there is not on the sincere use of "sweet," but this does not mean the terms express different properties or that otherwise identical sentences in which the terms are used express different propositions. The proposition expressed by an assertion that sugar is "tweet" is the same proposition as is expressed by the assertion that sugar is "sweet." This proposition obviously is not normative, and to me it seems implausible that our assertion should count as normative simply because of the special motivational sincerity condition on the use of the word "tweet." Since what holds in this imaginary case is parallel to what holds in the case of moral assertion, if motivational internalism is correct, I do not think that motivational internalism is plausibly viewed as an account of a kind of normativity.

Troubled by the artificiality of the example, one might reply that there is no *property* of tweetness. There is only sweetness and the contingent fact that we are motivated to eat sweet things. There *is*, however, the property of being morally required, and this property is such that, *necessarily*, a person who believes he is morally required to do something is motivated to do it. This reply raises difficult issues. Perhaps the property expressed

by the term “morally required” is such that it is merely contingent that people are motivated to do what they take to have that property *except that* they do not count as *conceiving* of an action as morally required unless they are motivated accordingly. That is, perhaps the parallel between the property expressed by “tweet” and that expressed by “morally required” is quite close, even though, contingently, the term in common use that expresses the property expressed by “tweet” is “sweet,” which does not implicate motivation, while the term “morally required” is in common use and, at least according to motivational internalists, it does implicate motivation.²²

The ‘tweet’ example shows two things, I believe. First, it shows that motivational internalism is compatible with naturalism. And second, it shows that motivational internalism is not a plausible account of a kind of normativity. Motivational *normativity* is relevantly different, however, for it is a thesis about the impact of moral beliefs on the motivation of *practically rational* persons. Hence, it seems to capture a normative aspect of moral belief that goes beyond mere generic normativity.²³

(3) *Authoritative Normativity*

It might seem that the ideas of generic normativity and motivational normativity are too weak to capture everything we have in mind in thinking of moral claims as normative. For example, even if moral claims are generically and motivationally normative, such that a rational agent would be motivated by her moral beliefs, it might be that she could experience this motivation as alien and unwelcome in something like the way that hunger

22 The suggestion in this paragraph is similar to a view developed in Tresan 2006. Tresan distinguishes between *de re* internalism and *de dicto* internalism. In this terminology, the view suggested in this paragraph is *de dicto* internalist.

23 Suppose that we stipulate that a person counts as believing something is *schpreet* if and only if she believes it is sweet and, in addition, unless she is not practically rational, she is motivated to some degree to eat it. Is it implausible in this case that the claim that something is *schpreet* is normative in the way that, I said in the text, it is implausible that the claim that something is tweet is normative? If so, then my argument perhaps goes too far. But it seems to me that, given my stipulations, the claim that something is *schpreet* *would* be normative. For suppose I am competent with the use of “*schpreet*” and I assert that sugar is *schpreet*. In this case, it seems, I have expressed the belief that sugar is sweet as well as the belief that I am motivated to eat sugar to the extent that I am practically rational. It is plausible, then, that what I have asserted is normative since it concerns what I will do if I am practically rational.

can be experienced as unwelcome when one is attempting to concentrate on an important matter. This would be compatible with the generic and motivational normativity of moral claims. In short, for a variety of reasons, it might seem that something beyond generic and motivational normativity must be cited in order to explain the normativity of morality. What might we have left out?

A variety of ideas could be proposed at this point, and it is somewhat artificial to talk as if there is one thing at issue, “authoritative normativity.” All that we have is the idea that there is a gap here, a place where something seems to be missing. This could be a mistake. It could be that morality has no further feature that can fill the gap we seem to detect.

The thought, however, is that morality has an objective authority over us, an authority that is not merely a matter of generic normativity, and an authority that is not reducible to the responses or motivations a rational person would have in virtue of her moral beliefs. The thought is that the *reason* that a rational person would take her moral beliefs appropriately into account in deciding how to live is that her moral beliefs concern something with objective authority over her. What could this come to? There seem to be two chief ideas in the literature.

First, objective authority is meant to explain *why* morality has *motivational* normativity. It is meant to explain why *rational* people pay attention to the demands of morality. They pay attention, and they are motivated appropriately, because, to the extent at least that they are fully rational and thinking clearly, they understand that morality is a source of authoritative *reasons*, reasons of a kind that any rational person would take into account, if she were aware of them, just in virtue of being rational. As Philippa Foot says, the idea is that moral considerations “necessarily give reasons for acting to any man,” or, as Stephen Darwall says, they are a source of reasons “with genuine deliberative weight.”²⁴ Call this the “authoritative reasons proposal.”

24 Foot *rejects* the idea that moral considerations “necessarily give reasons for acting to any man,” and in so doing, she in effect denies that morality has authoritative normativity (1978, 161). She thinks that if moral considerations do not “give reasons for acting to any man,” then, even if we say they give us “moral reasons,” these would not be robust enough to be genuine (p. 168, n. 8). Referring to Foot’s discussion, Darwall suggests that the idea is that moral considerations give us reasons “with genuine deliberative weight.” He agrees with her that so-called reasons of etiquette “are not guaranteed to be reasons unqualifiedly.” The same would be true of moral reasons, he suggests, if morality lacks authoritative normativity. See Darwall 1997, 306.

Second, objective authority is meant to answer the traditional challenge to morality posed by the question, Why be moral? It is taken to provide the definitive answer to the challenge.²⁵ Any putative requirement or ideal can be questioned. We can ask, Why pay attention to etiquette? Perhaps we can even ask, Why be rational? But if a requirement or ideal has authoritative normativity, then, the idea is, a rational person who was thinking clearly would understand that the relevant why-question is closed. That is, the question could not arise as a genuine practical question for any rational person who was thinking clearly about, and who understood the nature of, a requirement or ideal that has authoritative normativity. Call this the “closed question proposal.”

These are different proposals, but they are compatible with each other. According to the authoritative reasons proposal, a person who was rational and thinking clearly would understand that morality is a source of authoritative *reasons*, reasons of a kind that any rational person would take into account if she were aware of them, just in virtue of being rational. Yet understanding that morality is a source of authoritative reasons might not close the Why be moral? question for her, for she might think there are nonmoral sources of authoritative reasons, and that such reasons could override the reasons deriving from morality. Hence, it seems, the closed question proposal is stronger than the authoritative reasons proposal. On the closed question proposal, a rational person who is thinking clearly and who understands the nature of morality thereby understands something about morality such that, for her, the Why be moral? question cannot arise as a live question. It might seem that what she must understand is that the authoritative reasons that derive from morality are not overridden by any other reasons.²⁶

Consider the myth of Gyges, as told by Plato in the *Republic*.²⁷ Gyges is a shepherd who discovers a ring that makes him magically invisible when he turns it on his finger. Using the ring, Gyges seduces the queen, kills the king, and takes over the kingdom. It seems fair to say that this course of behavior was morally unacceptable. Now suppose that Gyges becomes persuaded of the wrongness of his plan in advance of the seduction. He might ask himself, “Why should I be moral?” especially given that doing

25 Darwall suggests that the issue whether morality is normative is posed by the “Why be moral?” question (2004).

26 But since there may be other ways of understanding the proposal, I do not want to build the thesis that moral reasons are overriding into the closed question proposal itself.

27 Plato 1992, 359d–360b.

his duty would mean giving up a life of power and love. Merely responding, “Because it’s your duty,” would not cut any ice since he already knows where his duty lies and he is troubled by the thought that he must follow through.

Consider, then, whether invoking any of the three grades of normativity would help Gyges with his question. Merely to invoke generic normativity would not address the question. Gyges presumably is aware that the *wrongness* of his plan has just the kind of relevance to his decision about what to do that makes it generically normative. This still leaves him with the question *why* he should avoid wrongdoing. And it would not help to invoke motivational normativity. As a rational person, Gyges could find himself motivated to do what he understands to be his duty, but he might still wonder *why* he should do his duty. Yet if morality has objective authority, then, the idea is, if Gyges understood the nature of moral considerations and if he were rational and thinking clearly, he would realize that the wrongness of his plan gives him an *authoritative* reason not to carry through with it. Furthermore, if he were rational and thinking clearly and if he understood the nature of moral considerations, his question about whether to be moral would not seem to be a live question for him.

One might think that, even given all of this, Gyges could still raise the question, “What are these authoritative reasons to me?” It does need to be said that the authoritative normativity of morality cannot silence every kind of doubt. Theoretical doubts and puzzles might still arise, including doubts about authoritative reasons. Yet the idea is that a *rational* person who *understands* the concept of objective authority and *believes* that moral requirements and ideals have objective authority cannot think there is a serious practical question about whether to be moral. Such a person might raise various theoretical questions, but she would not be uncertain or indecisive about whether to be moral, not if she were fully rational and thinking clearly.

This completes my explication of the three grades of normativity. Of course, I distinguished two proposals about authoritative normativity, and there may be other ideas as well. There is, therefore, some reason to think it would be useful to distinguish more than three grades of normativity. But this is how I will organize the discussion.

4. THE PROBLEM REVISITED

If we think that morality does have these three grades of normativity, the next question to ask would be, Can a moral *naturalist* explain or account

for this? But a naturalist might want to deny that one or more of the grades of normativity is genuinely characteristic of morality. For each of the three grades of normativity, then, there is a pair of questions. How plausible is it that morality is normative in this way? And, can a naturalist account for this kind of normativity?

In order to argue against naturalism on the basis of the normativity of morality, therefore, one would have to defend two key premises. The first premise would be to the effect that morality is normative on some given understanding of what this amounts to. And the second premise would be to the effect that moral naturalism cannot accommodate the fact that morality is normative on this understanding. Either of these premises can be controversial. If we let “X-normativity” refer to one of the three grades of normativity, or else to some other property that is a candidate for a kind of normativity, the naturalist might object that morality does not have the property of being X-normative. Alternatively, she might argue that naturalism can account for X-normativity. Her opponent may face some unexpected complications.

First, in order to undermine moral naturalism without undermining moral realism, the anti-naturalist must argue that the difficulty in accounting for X-normativity can be traced specifically to the naturalist’s idea that the moral properties are natural. Otherwise, it might appear that what the argument really does is to undermine the realist thesis that there are moral properties. Moreover, second, a successful anti-naturalist argument must rule out the strategies I mentioned before that seek to explain normativity without treating it as a property of the moral properties. Otherwise a naturalist might be able to explain X-normativity using one of these strategies.

From this point on, I will discuss the three grades of normativity, considering, first, the plausibility of the idea that morality has that grade of normativity, and second, the plausibility of the idea that moral naturalism is somehow incompatible with the hypothesis that morality has that grade of normativity. But a brief digression is in order.

5. EXPRESSIVISM AND REALIST-EXPRESSIVISM

It may seem that I have missed a well-known and obvious account of normativity as well as an obvious argument against moral realism. One of the realist doctrines is the view, often called ‘descriptivism,’ that a primary semantic role of the moral predicates is to ascribe moral

properties. A number of thinkers, from Charles Stevenson and R. M. Hare through Simon Blackburn and Allan Gibbard, have shared the view that ‘descriptivist’ accounts of moral thought and discourse cannot account for normativity.²⁸ Their idea is that normative thoughts involve conative attitudes of one kind or another and that descriptivist views cannot satisfactorily account for this. Gibbard claims, for instance, that “the special element that makes normative thought and language normative” is that “it involves a kind of endorsement – an endorsement that any descriptivistic analysis treats inadequately.”²⁹ If successful, this objection would cut equally against all versions of moral realism, whether naturalist or nonnaturalist.

However, the objection is not successful. Consider, for example, that we would be expressing an attitude of contempt in saying that cursing is typical among “rednecks.” Even so, it is clear that one of the central semantic roles of the term “redneck” is to ascribe the property of, roughly, being “a member,” perhaps an uneducated or “uncouth” member, “of the white rural laboring class.”³⁰ The example illustrates how it could be the case both that we use moral terms, such as “wrong” and “good,” to express attitudes such as disapproval and approval, and that a central semantic role of such terms is to ascribe moral properties. The expressive speech act potential of the terms does not undercut the realist thesis that the terms are used to ascribe moral properties. Expressivists are therefore mistaken to think that moral naturalism cannot account for the fact – if it is a fact – that moral thought and discourse involve having or expressing conative attitudes of one kind or another. The naturalist can say that moral discourse and thought involve a kind of endorsement and that they also ascribe moral properties. I have explained elsewhere in some detail how such an account might go. I have called this kind of view “realist-expressivism.” It is realist – it supposes there are moral properties – and it is expressivist – it supposes that moral thought and discourse involve “a kind of endorsement.”³¹

The example of the term “redneck” suggests, however, the implausibility of the view that, in Gibbard’s words, “the special element that

28 Stevenson 1937; Hare 1952; Blackburn 1993; Gibbard 1990.

29 Gibbard 1990, 33.

30 Sometimes it appears that the extension of the term is restricted to white people living in rural areas of the southern United States. I quote from the *American Heritage Dictionary of the English Language*, 3rd ed., s.v. “redneck.”

31 Chapter 5.

makes normative thought and language normative” is that “it involves a kind of endorsement.” “Redneck” is not a normative term and thought about rednecks is not normative. So it seems to me that even if moral thought and language involve endorsement and similar conative attitudes, it is not plausible that this is sufficient to account for their normativity. It seems that the *subject matter* of moral thought and discourse is normative. Arguably, this means that the normativity of moral thought and discourse is due to a feature of the moral *properties* that we attribute to things in having moral thoughts rather than being due merely to the expressive force of moral predicates.³² Even if realist-expressivism is part of a fully adequate naturalism, it is not sufficient to tell the full story.³³

This brings us to the central issue, which is whether *natural* properties can be *normative*. This is the territory on which the central dispute between ethical naturalism and nonnaturalism is to be found. There are the three grades of normativity to consider.

6. GENERIC NORMATIVITY AND MOTIVATIONAL NORMATIVITY

It is surely correct that moral thought and discourse are characterized by *generic* normativity. The idea here is that, for example, the claim that lying is wrong is evaluative or proscriptive in virtue of what it says or in virtue of its content. It is that, as I said before, moral claims have a characteristic essential semantic connection to decisions about action or choice. If we compare the claim that lying is wrong with the claim that lying is widespread, a naturalist will say that different properties are at issue, the property of wrongness and the property of being widespread, respectively. From the perspective of any moral realist, this will seem to be the key difference between the contents of the claims, leaving aside the issue of expressive content. This means that if we are to explain generic normativity by reference to the content of moral claims, it seems we must explain it by reference to a characteristic normativity of moral properties.

One might think that this is enough to settle the matter against naturalism since one might have the picture of properties as just being there

32 Jonathan Dancy makes a similar claim (2006, 136–138).

33 In chapter 5, I argue that realist-expressivism can explain why motivational internalism or judgment internalism can appear to be true even though it is not true.

in things, inertly, with no intrinsic bearing on action or choice. But this would be a mistake. Of course, it would cut against any kind of moral realism, whether naturalist or nonnaturalist. And it would be question begging, for moral realism is precisely the view that there are such properties as wrongness, and no property would count as *wrongness* if it did not have a bearing of the relevant kind on action. This is just to repeat the insight that wrongness and other moral properties must have generic normativity. It is true, however, that a moral realist cannot get away with saying merely this. We are owed an explanation of what the property of generic normativity could amount to.³⁴

I have proposed such an explanation in other work, where I have developed a naturalistic account of generic normativity called the “standard-based” account.³⁵ The idea is that a normative property relates an action or object of choice to a norm or standard with a relevant status. For example, I argued, the property actions can have of being rude, or being ruled out by (the local) etiquette, is a property that relates actions to norms or standards that are shared in the familiar way in the local culture. And I argued that moral wrongness is, roughly, the property of being ruled out by a system of standards the currency of which in the relevant society actually would contribute better than that of any other such system to enabling the society to meet its needs. When generalized in an appropriate way, this theory provides an account of the other moral properties, such as the property of being morally required, the property of being supported by a moral reason, and so on. This approach could be disputed, but it is a naturalistic account.

Motivational normativity is more controversial. I have argued in other work that it is possible for a rational agent to believe that she is morally required to do something yet not to be motivated to do it. Indeed some counter-examples to motivational *internalism* seem also to be counter-examples to motivational *normativity*. Depression can remove a person’s motivation to do what she believes she is morally required to do, and it seems to me that a depressed person is not necessarily irrational in any ordinary sense. A depressed person might think it is more important for

34 I made this complaint – although not in these terms – in two earlier papers on moral naturalism (Copp 1990 and 1991).

35 See Copp 1995, esp. ch. 1 and ch. 11, pp. 223–231. For a brief exposition, see Copp 1996 and Copp 1997b. For a reply to some objections, see Copp 1998. See also the introduction in this book.

the moment to think about her lost loves than to be fastidious about what she is morally required to do, and so she might have no motivation to do something that she believes she ought to do. This might not be any indication of irrationality.³⁶ I therefore find it doubtful that morality is characterized by motivational normativity.³⁷

When a person fails to do what she believes she is morally required to do, having lacked *any* motivation to do it, she is presumably blameworthy. Of course we think that a person can be excused in some cases for failing to do her duty, but we do not think that a person can be excused for lacking *any* motivation *at all* to do what she believes she is morally required to do. This said, however, we would not likely think that such a person must be irrational, as would be implied by the thesis of motivational normativity. Intuitively, her lack of moral motivation reflects on her moral character, not on her rationality. That is, the thesis of motivational normativity seems counter-intuitive, and not at all a matter of ordinary thinking about morality.

All of this being said, however, it is important to point out that moral naturalism can accommodate the idea that morality has motivational normativity. For, in the first place, it is plausible that a rational person is disposed to be moved by reasons. More precisely, it is plausible that a rational person is so disposed that, if she took herself to have a reason to do something, she would be motivated to some degree to do it, at least under certain circumstances. Perhaps she would not be so motivated if she thought there were stronger reasons against doing the thing than in favor, but in the absence of an awareness of any countervailing reasons, it is plausible that she would be motivated to do the thing. A naturalist can hold that such a disposition is partly constitutive of rational agency. In the second place, it is plausible that moral considerations are a source of reasons. Indeed, on my account of generic normativity, if a person is morally required to do something, then there is a moral reason for her to do it. A naturalist might then hold that a rational person is so disposed that, if she takes herself to be morally required to do something, and if she understands what this involves and is thinking clearly, then she would be motivated to some degree to do the thing, at least under certain

36 Here I follow arguments in Copp 1997a.

37 It is possible that realist-expressivism can help to explain why motivational normativity might seem to be true in something like the way it helps to explain why motivational internalism can appear to be true.

circumstances. This position is compatible with moral naturalism, but I think it is mistaken.

I agree, of course, that it is constitutive of rational agency to be so disposed that, if one takes oneself to have an *authoritative* reason to do something, one is motivated to some degree to do it, at least under certain circumstances. This follows from the concept of an authoritative reason, since authoritative reasons are, by stipulation, reasons of a kind that a rational person would be responsive to, if she were aware of them, just in virtue of being rational. Nothing here is incompatible with moral naturalism. But I think it is an oversimplification to suppose that rational persons are disposed to be moved by reasons of *every* kind. There are reasons of different kinds, and not all of them are authoritative. It is not a failure of rationality to lack any tendency to be moved by reasons of etiquette. More to the point, I think that moral reasons are not in general reasons that a rational person would be responsive to, if she were aware of them, just in virtue of being rational. That is, I believe that moral reasons are not authoritative reasons.

Philosophers who hold that morality has motivational normativity presumably would disagree. They might say that moral reasons are not *genuine* reasons if they are not among the reasons that a rational person would be responsive to, if she were aware of them, just in virtue of being rational. This strikes me as an eccentric use of the word “genuine,” but the substantive issue is whether moral reasons are genuine in this sense. My claim here is simply that there is logical space to accept that *there are* moral reasons while rejecting motivational normativity.³⁸

It appears, then, that if normativity makes difficulty for moral naturalism, the difficulty lies with the third grade of normativity, authoritative normativity. To be sure, critics will argue that the naturalistic accounts of generic and motivational normativity that I have sketched are unsuccessful. But it would be more difficult to show that *no* naturalistic account *could* be successful. In addition, critics would have to defeat skeptical arguments, such as the argument I offered above, against the thesis that morality is characterized by motivational normativity. The better strategy for critics of naturalism, it seems to me, would be to argue that generic and motivational normativity are insufficient to account for the special authority that morality has, and then to argue that no naturalistic theory could explain authoritative normativity.

38 Cf. Foot 1978, esp. 160–162 and 168 n. 8.

7. AUTHORITATIVE NORMATIVITY

There is a widely shared intuition that generic and motivational normativity are insufficient to explain the normativity of morality. To be sure, if I am correct, morality does not actually have motivational normativity. Yet it might seem that the examples I have pointed to, in which a rational person fails to be motivated appropriately by her moral beliefs, are compatible with an underlying conceptual link between moral belief and motivation, a link that depends on and is explained by the authoritative normativity of morality.

As discussed before, there are two proposals about this. According to the first, moral considerations give us authoritative reasons, reasons of a kind that any rational person would take into account, if she were aware of them, just in virtue of being rational. More generally, requirements of some kind *K* have authoritative normativity just in case, necessarily, there is an authoritative reason to act as one is *K*-required to act – a reason that any rational person would take into account, if she were aware of it, just in virtue of being rational. According to the second proposal, a rational person who was thinking clearly would understand that the “Why be moral?” question is closed. That is, the question could not arise as a live question for her, a question that indicated any indecision about whether to be moral. More generally, requirements of kind *K* have authoritative normativity just in case, necessarily, no fully rational person who is thinking clearly can take there to be any serious practical question of whether to act in accord with *K* requirements.

My discussion of authoritative normativity will be complex. First, I will raise the skeptical worry that, at least as they stand, neither the authoritative reasons proposal nor the closed question proposal is plausible as an account of a heightened grade of normativity. Basically the same worry also applies to motivational normativity. Second, I will point out that there are naturalistic strategies for explaining authoritative normativity, at least on the authoritative reasons proposal. Third, I will consider, very briefly, a family of ‘Kantian’ accounts of authoritative normativity that turn on a strategy of argument I call the “self-conception strategy.” I will use Christine Korsgaard’s views in order to illustrate how a naturalist might try to turn a self-conception theory to her advantage. Fourth, and finally, I will contend that it is not plausible that morality meets the conditions proposed by either the authoritative reasons proposal or the closed question proposal. It appears that morality does not have authoritative normativity. I begin with the skeptical worry.

(1) *Doubts about the Significance of Authoritative
and Motivational Normativity*

The idea of rational agency is crucial to the accounts I have given of both motivational and authoritative normativity. Motivational normativity is explained in terms of what would motivate a rational person. Authoritative normativity, on the authoritative reasons proposal, is explained in terms of reasons that a rational person would take into account in her deliberation; on the closed question proposal, it is explained in terms of a key question that no rational person would take to be an open practical question. The idea of rational agency that is invoked in these accounts is the idea of an agent who complies with norms of practical rationality, such as the norm that requires one to take the means to one's ends. That is, motivational and authoritative normativity relate anything that has them to the norms of rationality. Hence, if morality has motivational normativity, or if it has authoritative normativity, its having this property consists in the obtaining of the relevant specified relation between it and the norms of rationality. This means that unless norms of rationality have special normative significance, neither motivational normativity nor authoritative normativity can plausibly be viewed as a special or heightened grade of normativity. Let me explain more fully.

Begin with the authoritative reasons proposal, which explains authoritative normativity by reference to so-called authoritative reasons. On the face of it, such reasons are just a *kind* of reason alongside other kinds of reason. They are reasons that any *rational* person would take into account in deliberation, if she were aware of them, just in virtue of being *rational*. Moral reasons are reasons that any *moral* person would take into account, if she were aware of them, just in virtue of being *moral*. It would be interesting and perhaps important if moral reasons were also authoritative reasons in the specified sense. But in the absence of a showing that the norms of rationality have a special kind of normative significance, to establish that moral reasons are also authoritative would not be to show that they have a special or heightened grade of *normativity*.

One might protest that authoritative reasons are not just another kind of reason. They are reasons that have "genuine deliberative weight," which so-called moral reasons and other reasons-of-a-kind, such as reasons of etiquette, need not have.³⁹ It is unclear how to understand this protest, however, except as a pointing out that, while reasons of etiquette are

39 See Darwall 1997, 306.

not reasons that any rational person would take into account, if she were aware of them, just in virtue of being rational, authoritative reasons are, by stipulation, reasons that any rational person would take into account, if she were aware of them, just in virtue of being rational. We might insist on reserving the term “reason” for authoritative reasons, but this would not accomplish anything substantive.

Note that, if there are any requirements of rationality, they give rise to authoritative reasons.⁴⁰ For if an action is rationally required, then any person who is aware that it is rationally required would take this fact into account in her practical deliberation to the extent that she is rational, just in virtue of being rational. This conceptual point implies that there is authoritative reason to comply with the requirements of rationality. There is a corresponding point one could make about moral requirements, for moral requirements are such that, necessarily, any person who is aware that an action is morally required would take this fact into account in her practical deliberation to the extent that she is moral, just in virtue of being moral. This point about moral requirements does not show that they have any *special* authority, and the similar point about requirements of rationality does not show that *they* have any special authority.

The point I am making might be obscured if we already believe, perhaps on independent grounds, that the requirements of rationality have special normative significance. But in the absence of an argument that norms of rationality have special normative significance, it seems that the fact that the requirements of rationality give rise to authoritative reasons does not show that they have any special status. Similarly, then, the fact, if it is a fact, that moral requirements give rise to authoritative reasons cannot show that they have any special status. In the absence of an argument that norms of rationality have special normative significance, it would not show morality to enjoy a special degree or kind of normativity. It therefore seems that, in order to support the authoritative reasons proposal as a conception of a special or a heightened degree of normativity, we need a showing that rationality has special normative significance.

Turn now to the closed question proposal, which explains authoritative normativity by reference to a question that no fully rational person who is thinking clearly can take to be a serious practical question. It might seem

40 I take it as given that requirements of rationality would have generic normativity and motivational normativity. They would have motivational normativity, for it is presumably a conceptual truth that any rational person who believes he is rationally required to do something will be motivated to some degree to do it.

that no rational person who was thinking clearly could take there to be a serious practical question whether to act in accord with the requirements of rationality. Yet again, it might also seem, by parity of reasoning, that no fully moral person could take there to be a genuine question whether to do her moral duty. But this would not show that moral requirements have any special status, and the corresponding point about requirements of rationality surely could not show by itself that they have a special status. If not, then it seems that even if no fully rational person who was thinking clearly could take there to be a serious practical question whether to act in accord with moral requirements, this by itself could not show that moral requirements have a special status. It therefore seems to me, again, that in order to support the closed question proposal as a conception of a special or a heightened degree of normativity, we need a showing that rationality has special normative significance.

It should now be clear that a parallel argument can be constructed with respect to motivational normativity. That is, in order to support the idea that motivational normativity is a special or a heightened degree of normativity, we need a showing that rationality has special normative significance. In short, the three proposed grades of normativity merely relate morality or moral properties to requirements or norms of rationality, and so, unless the latter have special normative significance, a showing that morality stands in the specified relation to them cannot amount to a showing that morality has a heightened degree or grade of normative stringency.

It may be, of course, that requirements of rationality *do* have special normative significance. Elsewhere, I have proposed an account of what I call “self-grounded” rationality. In so doing, I argued that the requirements of self-grounded rationality, and the associated reasons, have a (defeasible) centrality in the deliberation of autonomous agents that other kinds of requirements and reasons need not have.⁴¹ It is not clear that this means they have a special *normative* significance. But if it does, this isn’t a problem for naturalism, for, although I cannot go into the details here, it can be explained in a way that is compatible with naturalism.⁴²

I have also argued elsewhere that requirements of self-grounded rationality do not have a special status at the level of generic normativity.⁴³ It

41 Chapter 10.

42 The argument in chapter 10 is entirely compatible with naturalism. Another naturalistic account is found in Hubin 2001.

43 Chapter 9.

now appears that any difference at the levels of motivational normativity or authoritative normativity between their status and the status of requirements of morality may be of no normative significance, since motivational and authoritative normativity are *defined* in terms of rational agency, which in turn is understood in terms of norms of rationality. We cannot use such differences to support the view that requirements of rationality have special normative significance, since these grades of normativity are defined in terms of rational agency, which *presupposes* the idea that norms of rationality have special significance. That is, the idea that motivational normativity or authoritative normativity is a heightened grade of normativity presupposes that norms of rationality have special normative significance.

(2) *Naturalism and Authoritative Normativity*

Let me set these worries aside, however. I now want to point out that there are naturalistic strategies for explaining authoritative normativity, at least on the authoritative reasons proposal, and perhaps also, as we will see in subsection 3, on the closed question proposal.

My account of “self-grounded” rationality is fully naturalistic. I cannot go into detail here,⁴⁴ but the central idea, very roughly, is that rationality consists in efficiently serving one’s values. I propose that rational action is governed by a standard I call the standard of self-grounded reason, which calls on us to serve our values. I argue that this standard is relevantly authoritative in virtue of the fact that conformity with it furthers or instantiates one’s self-government. It is in virtue of this fact that claims about the requirements of self-grounded reason have generic normativity. For present purposes, however, the important point is that if rationality consists in efficiently serving one’s values, then self-grounded reasons *are* the reasons that a rational person would take into account in deliberation if she were aware of them, just in virtue of being rational. If I am correct, it follows that self-grounded reasons are *authoritative*, on the above definition of authoritative reasons. This means, if I am correct, that naturalism can make room for authoritative reasons.

The interesting question that remains, of course, is whether *moral* reasons are authoritative. I would argue that a rational agent need not have any moral values, and that if she does not, she may not have self-grounded reason to do what she has moral reason to do. That is, a rational person

44 See chapter 10.

would not necessarily take moral reasons into account in deliberation if she were aware of them, just in virtue of being rational. Moral reasons are not authoritative reasons.

Michael Smith disagrees. He has developed a naturalistic theory according to which, in my terminology, moral reasons are authoritative. According to Smith, roughly, the moral rightness of an action in conditions C is the feature we would want acts to have in C if we were fully rational. He claims that this feature is a natural feature of acts, at least in worlds like ours.⁴⁵ The argument for this depends on two theses. First, the rightness of an action is a matter of its being called for by reasons, and second, there is a reason for an agent to do something just in case, roughly, if she were fully rational, she would want herself (as she actually is) to do it. Smith provides a broadly naturalistic account of full rationality in terms of the coherence of a rational person's beliefs and desires. Given all of this, he argues, to believe that one is morally required to do something is, roughly, to believe that, if one's beliefs and desires were fully coherent, one would want oneself (as one actually is) to do it. Moreover, if a fully rational person believes such a thing regarding an action she could perform in a situation, she must be motivated to perform that action. Moral reasons therefore qualify as authoritative on Smith's account, for they are reasons that would motivate a fully rational person just in virtue of her rationality. I have argued elsewhere that Smith's argument is unsuccessful,⁴⁶ but the important point is that the problem with Smith's account is *not* that it is *naturalistic*. It fails for other reasons.

(3) *Self-Conception Strategies*

A family of 'Kantian' views agrees with Smith that moral reasons are authoritative. Proponents of these views employ the 'self-conception strategy' to show that moral reasons are authoritative. On the self-conception strategy, there is a way of conceiving of oneself such that a rational person who is thinking clearly *must* conceive of herself in this way, but if she does not treat moral reasons as authoritative, she cannot *coherently* conceive of herself in this way. It might be said, for example, that a person who does not treat moral reasons as authoritative cannot see herself as *autonomous*; or that she cannot see herself or value herself as a rationally reflective *agent*,

45 See Smith 1994, esp. chs. 5 and 6; on the identification of rightness with a natural feature of acts, see pp. 184–186; on the definition of naturalism, see p. 203 n. 1.

46 See Copp 1997a.

acting for *reasons*; or that she is committed to a *practical solipsism*; or that she cannot coherently expect *other* people to respond to the reasons she addresses to them; or the like.⁴⁷ On the self-conception strategy, rational people who are thinking clearly and coherently respond appropriately to moral reasons.

Given the complexity and variety of self-conception views, there is little hope of a generic assessment of them. They need to be evaluated one at a time. They make two central claims, each of which needs defense. First, they specify some way of understanding ourselves such that, they claim, rational people who understand morality and are thinking clearly will not be able to understand themselves in this way if they are unmoved by moral considerations. Second, they claim that the specified way of understanding ourselves is non-optional, such that no rational person who was thinking clearly could fail to view herself in this way. It is only if we cannot fail to view ourselves in the relevant way, provided that we are rational and thinking clearly, that it follows from the first claim of a self-conception view that we cannot fail to be moved by moral considerations provided that we are rational and thinking clearly.

For my purposes, the key point is that a self-conception view may be compatible with moral naturalism. To illustrate this, I want to look briefly at Christine Korsgaard's position. The central point is that one cannot use a self-conception view to argue against moral naturalism without showing that the view is incompatible with naturalism.

Korsgaard says that the central question that needs to be answered is, "[W]hat *justifies* the claims that morality makes on us?" She calls this the "normative question."⁴⁸ She suggests that whatever property an action might have, a person can intelligibly ask why that should matter to her, or why she should take that to be a reason for her to act. So it appears we cannot answer the normative question by citing a property of an action, and she argues that an action's being morally required is not a matter of its having a certain kind of property.⁴⁹ Instead, she holds, an action's

47 I sketch these ideas with apologies, respectively, to the following philosophers: Immanuel Kant (1785), Christine Korsgaard (1996), Thomas Nagel (1970), and Stephen Darwall (2006).

48 Korsgaard 1996, 9–10.

49 Ibid., 33–35. She therefore rejects the position she calls "substantive realism." The substantive realist postulates the existence of "intrinsically normative entities" such that correct answers to moral questions are correct "*because* there are moral facts or truths [involving those entities], which those questions ask *about*" (ibid., 34–35, emphasis in original).

being morally required is a matter – very roughly – of the action's being required by our "identity" as rationally reflective agents – where a person's identity is the set of properties she values in herself.⁵⁰ Reflective agency, she thinks, is a property that we cannot but value having, if we are to value anything. Accordingly, she thinks, the requirements of our identity as rationally reflective agents are not open to the normative question.

Korsgaard's normative question is very much like the traditional "Why be moral?" question, and indeed, in some passages she seems to be asking this question.⁵¹ It appears that, on Korsgaard's account, the "Why be moral?" question could not arise as a genuine practical question for a fully rational person who was thinking clearly and who understood what it is for an action to be morally required – provided she understood in addition that she must value her nature as a rationally reflective agent if she is to value anything. Such a person, Korsgaard thinks, does not view the normative question, "What justifies the demands that morality makes on us?" as raising a genuine practical concern. And if not, then to the extent that she is fully rational and thinking clearly, the "Why be moral?" question also does not raise a genuine practical concern.

I am not persuaded. I think that a rational person who understood Korsgaard's account might still ask why she should be moral, and this question might, for her, be a practical question, a question that indicates an indecisiveness about whether to be moral. Gyges values the power and love he will achieve if he carries out his plot, and this means that, if he is fully rational and thinking clearly, and if Korsgaard's theory is correct, he must value his reflective agency. Moreover, if Korsgaard's analysis is correct, and if Gyges understands morality, he must then understand that carrying out his plot would conflict with his valuing his reflective agency. Yet he also values power and love, and he understands that carrying out his plot would help him to achieve a life of power and love. So he might ask, "Why should I not carry out my plot? Why should I be moral?" For all that Korsgaard has shown, it seems to me, his asking these questions would not indicate either that he is not fully rational or that he does not understand morality. Nor need it indicate that he does not value his reflective

Although Korsgaard rejects "substantive realism," she accepts "procedural realism," which she defines as the view "that there are right and wrong ways to answer [moral questions]" (ibid., 35). She adds, "Substantive realism is a version of procedural realism, of course" (ibid., 37 n. 58).

50 Ibid., 101.

51 Ibid., 33.

agency. A person who values A can rationally wonder whether to do something that he understands would conflict with achieving or sustaining A when doing it would allow him to achieve something *else* he values. To wonder about this is compatible with his continuing to value A.⁵²

The more important point, however, is that a naturalist can embrace a reconstruction of Korsgaard's position. Suppose that Korsgaard is correct that if an action is required of a person, then the person's not performing it would conflict with her valuing herself as a rationally reflective agent. Suppose she is also correct that a person cannot but value her rational reflective agency unless she ceases to value anything, and thereby ceases to be able to see herself as an agent. All of this specifies a complex property that, if Korsgaard is correct, is possessed by actions that are morally required – the property an action has if and only if, roughly, (K1) the agent's failure to perform it would conflict with her valuing herself as a rationally reflective agent and (K2) the agent must value rational reflective agency if she is to value anything and to see herself as an agent. Call this the "Korsgaard property." I say that it could be viewed as a natural property.

How could the Korsgaard property be a natural property? Earlier I suggested that a property is natural just in case, roughly – leaving aside conceptual truths – any proposition about the instantiation of the property that can be known can be known only empirically. Consider, then, propositions to the effect that an action has the Korsgaard property. Any such proposition is equivalent to the conjunction of a 'K1 proposition,' a proposition to the effect that the action in question has property K1, with a 'K2 proposition,' a proposition to the effect that the action has property K2. Therefore, we can divide our problem in two. Begin with K2. A naturalist who is tempted by Korsgaard's picture presumably needs to argue that it is a conceptual truth that any agent must value rational reflective agency if she is to value anything at all and to see herself as an agent. But if this is a conceptual truth, then it is also a conceptual truth that any action has the K2 property. (Similarly, if it is a conceptual truth that murder is

52 My argument here depends on an interpretation of the idea that actions may be *required* by our *identity* as rationally reflective agents. I assume this means, for Korsgaard, that actions may be required by our valuing rational reflective agency. And I propose that an action is required by our valuing something just in case a rational and factually informed person who valued that thing and who was thinking clearly could not fail to perform the action *provided* that nothing else he valued was at stake. One might respond that the final clause of my proposal should be deleted. But if I am correct, deleting this clause would mean in effect that Gyges is not required by his valuing rational reflective agency to abandon his plot. For he faces a conflict of values. He can choose to pursue power and love while continuing to value rational reflective agency. See Copp 1999.

wrong, then it is a conceptual truth that any action has the property of being such that murder is wrong.) But if it is a conceptual truth that any action has the K2 property, then the relevant K2 propositions do not raise a problem for the naturalist. Turn now to K1. A naturalist who is tempted by Korsgaard's picture presumably must argue that, for some types of action, it is a conceptual matter, and for other types of action, it is an empirical matter, that actions of that type have the K1 property – the property of being such that an agent's failure to perform an action of that type would conflict with her valuing herself as a rationally reflective agent. Perhaps, for example, it is a conceptual matter that actions that are fair have the K1 property, and perhaps it is an empirical matter that donating money to charity has the K1 property. If the naturalist can make a case for claims of this kind about the K1 and K2 properties, then she can make a case that Korsgaardian rightness is a natural property. In short, if Korsgaard is correct, then an action is right just in case the relevant K1 and K2 propositions are true. If the naturalist can argue that such propositions are either empirical or conceptual, then she can accept Korsgaard's position.

(4) *That Morality Lacks Authoritative Normativity*

As I have suggested, I do not believe that moral reasons are authoritative in a way that puts an end to all doubts about whether to be moral. I said before that it might seem that a rational person could not question whether to live in accord with the requirements of rationality. But this is not the case, and it will be instructive to see why. Consider again the myth of Gyges. Let us suppose that Gyges thinks that in view of his values he ought *rationally* to carry on with his plan to take over the kingdom but that he ought *morally* to give up his plan. We imagined him as asking, "Why should I be moral?" and this seemed an intelligible question for him to ask since doing his duty would mean giving up a life of power and love. But we can also imagine him asking, "Why should I be rational?" and this also seems an intelligible question for him to ask since pursuing his plan would involve him in murder, and it would therefore mean violating his moral duty. It would be a mistake to suppose that, if he is rational and thinking clearly, he will settle on carrying out his plan, given that he understands that he rationally ought to pursue his plan. He might well understand that his reasons to pursue his plan have "genuine deliberative weight," such that he will take them into account provided that he is rational. But his question is whether to continue to be rational. It seems to me that he could decide against this. And if Gyges can take there to

be a genuine practical question whether to be rational, compatibly with his actually being rational and thinking clearly, he can surely take there to be a genuine practical question whether to be moral, compatibly with his being rational and thinking clearly.⁵³

I think in fact that a naturalist can reject both the authoritative reasons proposal and the closed question proposal, for I think it is a mistake to suppose that morality has authoritative normativity on either of these proposals.⁵⁴ In my view, as I suggested earlier, moral reasons are not reasons of a kind that any rational person would take into account in her deliberation, if she were aware of them, just in virtue of being rational. I have argued this in other places.⁵⁵ Moreover, there is no definitive answer to the “Why be moral?” question of the kind that believers in authoritative normativity have in mind. Of course there are answers that can be given to various individuals in their specific concrete situations. And there is a very general answer that can be given. For in my view, what justifies the demands of morality is that they are the demands imposed on us by a system of standards the currency of which in the relevant society would contribute better than that of any other such system to enabling the society to meet its needs. But this proposal does not yield a definitive answer to the “Why be moral?” question, for one can respond to my account by asking, “What is all of this to me?” There is no way to put all such doubts to rest in rational reflective persons. In my view, there is no answer to the “Why be moral?” question that lays it to rest by showing that it does not raise

53 To take an extreme case, suppose that the evil genius threatens to destroy the earth unless you take a pill that will make you no longer rational. Even though you are rational, you might decide to take the pill. Thomas Schelling has said that, in face of a threat, it is not “invariably an advantage to be rational” (1960, 18–19). We can imagine as well a case in which the evil genius threatens to destroy the earth unless you take a pill that will make you no longer be moral. Even if you are moral and rational, you might take the pill.

54 Indeed, it may seem that I have already argued that it is a mistake in my 1997 essay, “The Ring of Gyges: Overridingness and the Unity of Reason,” which is chapter 9 in this book. That chapter addresses the idea that moral requirements are “overriding,” and argues, in effect, that no account of overridingness can be given solely on the basis of an account of generic normativity. The issue is different, however, from the issue whether moral requirements have authoritative normativity. My arguments in chapter 9 seem to leave open the possibility that moral reasons are also reasons that any rational person would take into account, if she were aware of them, just in virtue of being rational. Even if moral reasons are authoritative in this sense, however, it does not follow that they are overriding. For one thing, there might be other kinds of authoritative reasons and moral reasons might not be normatively more important than they are.

55 See chapters 9 and 10.

a serious practical question that could indicate an indecisiveness about morality in a rational person who had a clear understanding of morality. If this is correct, morality does not have authoritative normativity – not, at least, on either of the conceptions of this that we have considered.

I do not think that a failure to be motivated by moral belief is a sign of irrationality. It is a sign of immorality, but this is no surprise. Nor do I think that a fully rational person who had a clear understanding of the nature of morality could not be indecisive about whether to be moral. Such a person could be indecisive about whether to continue to be rational, so it would be astonishing if she could not be indecisive about morality. Perhaps a fully *moral* person would not be indecisive about whether to be moral, but why should we expect *rationality* to secure *moral* motivation in a clear thinking and fully informed person? Those who think that morality has authoritative normativity therefore face a challenge. They need to provide an adequate conception of authoritative normativity and an argument to show that morality possesses authoritative normativity. And, if they think this counts against naturalism, they need to provide an argument that it does.

8. CONCLUSION

What I have tried to do in this chapter is to point to some of the complexities that would have to be addressed before we could conclude that moral naturalism is ruled out by the fact that morality is normative. I distinguished three grades of normativity: generic, motivational, and authoritative normativity. I argued that moral naturalism has no obvious difficulty explaining generic and motivational normativity. There are naturalistic theories in the literature, including a theory I have proposed, that purport to explain what is involved in these grades of normativity. Authoritative normativity is another matter. I discussed two proposals about what authoritative normativity consists in, and argued that morality lacks authoritative normativity on both proposals. I argued as well that, on either proposal, to support the idea that authoritative normativity is a special or a heightened degree of normativity, we need a showing that rationality has special normative significance. I argued that there are nevertheless naturalistic accounts of authoritative normativity on both proposals. I myself suspect that morality does not possess authoritative normativity. If I am correct about this, then it is no problem for naturalism if, despite what I have suggested, authoritative normativity cannot be understood naturalistically.

Let me conclude by saying that moral nonnaturalism faces the challenge of explaining the normativity of morality just as much as does moral naturalism. If normativity needs to be explained, it is not explained by giving up on naturalistic ways of explaining it. Antireductionist forms of nonnaturalism that view moral properties as *sui generis* face an especially difficult problem, for they appear simply to postulate normativity. It is unclear how they could explain it. If one of our goals is to explain what normativity consists in, then, as I have suggested, naturalistic accounts can be given of all three grades of normativity. These accounts seek to explain the normativity of moral properties by providing an analysis, first, that explains what normativity consists in, and, second, that exhibits normativity as a second-order natural property of the moral properties. Moral naturalism appears to be alive and in good health.

REFERENCES

- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- Copp, David. 1990. "Explanation and Justification in Ethics." *Ethics* 100: 237–258.
- . 1991. "Moral Realism: Facts and Norms." *Ethics* 101: 610–624.
- . 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 1996. "Moral Knowledge in Society-Centered Moral Theory." In *Moral Knowledge? New Readings in Moral Epistemology*, ed. Walter Sinnott-Armstrong and Mark Timmons, 243–266. New York: Oxford University Press.
- . 1997a. "Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*." *Ethics* 108: 33–54.
- . 1997b. "Does Moral Theory Need the Concept of Society?" *Analyse und Kritik* 19: 189–212.
- . 1998. "Morality and Society – The True and the Nasty: Reply to Leist." *Analyse und Kritik* 20: 30–45.
- . 1999. "Korsgaard on Rationality, Identity, and the Grounds of Normativity." In *Rationality, Realism, Revision*, ed. Julian Nida-Rümelin, 572–581. Berlin: de Gruyter.
- , ed. 2006. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
- Dancy, Jonathan. 2006. "Nonnaturalism." In Copp 2006: 122–145.
- Darwall, Stephen. 1983. *Impartial Reason*. Ithaca, N.Y.: Cornell University Press.
- . 1997. "Reasons, Motives, and the Demands of Morality: An Introduction." In *Moral Discourse and Practice*, ed. Stephen Darwall, Alan Gibbard, and Peter Railton, 305–312. New York: Oxford University Press.
- . 2004. "Normativity." In *Routledge Encyclopedia of Philosophy Online*, version 2.0, ed. Edward Craig. [Http://www.rep.routledge.com](http://www.rep.routledge.com).
- . 2006. "Morality and Practical Reason: A Kantian Approach." In Copp 2006: 282–320.

- Foot, Philippa. 1978. "Morality as a System of Hypothetical Imperatives." In *Virtues and Vices*, 157–174. Berkeley: University of California Press.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hubin, Donald C. 2001. "The Groundless Normativity of Instrumental Rationality." *Journal of Philosophy* 98: 445–468.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Kant, Immanuel. 1785. *Grounding of the Metaphysics of Morals*. Trans. James W. Ellington. Rpt., Indianapolis, Ind.: Hackett, 1981.
- King, Jeffrey C. 1998. "What Is a Philosophical Analysis?" *Philosophical Studies* 90: 155–179.
- Korsgaard, Christine. 1986. "Skepticism about Practical Reason." *Journal of Philosophy* 83: 5–25.
- . 1996. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge: Cambridge University Press.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth, England: Penguin.
- Moore, G. E. 1903. *Principia Ethica*. Ed. Thomas Baldwin. Rpt., Cambridge: Cambridge University Press, 1993.
- . 1968. "Reply to My Critics." In *The Philosophy of G. E. Moore*, 3rd ed., ed. P. A. Schilpp, 535–687n. La Salle, Ill.: Open Court.
- . 1993. "Preface to the Second Edition." In *Principia ethica*, ed. Thomas Baldwin, 1–37. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press.
- Plato. 1992. *Republic*. Trans. G. M. A. Grube and C. D. C. Reeve. Indianapolis, Ind.: Hackett.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. New York: Oxford University Press.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Stevenson, Charles. 1937. "The Emotive Meaning of Ethical Terms." *Mind* 46: 14–37.
- Sturgeon, Nicholas L. 2006. "Ethical Naturalism." In *Copp* 2006: 91–121.
- Tresan, Jon. 2006. "De Dicto Internalist Cognitivism." *Nous* 40: 143–165.

The Ring of Gyges: Overridingness and the Unity of Reason

1. INTRODUCTION

Does morality override self-interest? Or does self-interest override morality? These questions become important in situations where there is conflict between the overall verdicts of morality and self-interest, situations where morality on balance requires an action that is contrary to our self-interest, or where considerations of self-interest on balance call for an action that is forbidden by morality. In situations of this kind, we want to know what we ought *simpliciter* to do. If one of these standpoints overrides the other, then there is a straightforward answer. We ought *simpliciter* to act on the verdict of the overriding standpoint.

For purposes of this chapter, I assume that there are possible cases in which the overall verdicts of morality and self-interest conflict. I will call cases of this kind “conflict cases.” The verdict of morality in a conflict case would be a proposition as to what we ought morally to do, or as to what we have the most moral reason to do; the verdict of self-interest would be a proposition as to what we ought to do in our self-interest, or as to what action is best supported by reasons or considerations of self-interest.¹ These propositions are action-guiding or normative in a familiar

I presented earlier versions of this chapter at the 1995 conference on “Self-Interest” sponsored by the Social Philosophy and Policy Center at Bowling Green State University and to the departments of philosophy at Indiana University, the University of Arizona, and the University of Toronto. I am grateful to everyone who participated in the discussions on these occasions as well as to Philip Clark, James Drier, Anil Gupta, Ishtiyaque Haji, Dale Jamieson, Michael Jubien, Jeffrey C. King, and George Nakhnikian for helpful comments and suggestions about the issues addressed in the chapter.

1 I do not assume that the overall verdicts of morality and self-interest are always that some particular action is required. For all that I say, quite different verdicts are possible, including

sense.² The conflict between morality and self-interest in conflict cases is therefore a normative conflict; it is a conflict between the overall verdicts of different normative standpoints. I take it that the question whether morality overrides self-interest is the question whether the verdicts of morality are *normatively more important* than the verdicts of self-interest. In due course, I will explain the idea of normative importance as well as the ideas of a normative proposition and of a reason.

I will be defending the position that neither morality nor self-interest overrides the other, that there simply are verdicts and reasons of these different kinds, and that there is never an overall verdict as to which action is required *simpliciter* in situations where moral reasons and reasons of self-interest conflict. Accordingly, I reject the position that, in each situation, all the reasons there are determine one overall verdict, the verdict we might call the verdict of “Reason” or “Reason-as-such.”³ In my view, there is no standpoint that can claim normative priority over all other normative standpoints and render a definitive verdict on the relative significance of moral and self-interested reasons. That is, in cases of conflict between kinds of reasons, there is no fact as to what a person ought *simpliciter* to do. I will explain these claims. I will be defending a kind of skepticism about the unity of practical reason.

From one point of view, it will seem that my position threatens the rational significance of morality, and its significance in guiding our actions. From this point of view, morality purports to be the final arbiter of how to act. Morality purports to tell us what we ought to do, period, and without qualification. If moral reasons do not override all others, however, then these appearances are illusory. For if moral reasons are not overriding, there might not be sufficient reason for a person to act morally in situations where the morally required action would be contrary to her self-interest. A person may not be guilty of any failure of rationality if she always acts in her self-interest in such situations.

From a second point of view, however, my position may seem to offer liberation from morality rather than to threaten to discredit it. From this standpoint, the demands of morality are sometimes excessive, for it is sometimes impossible to comply with these demands while also living

the verdict that a situation is a moral dilemma. For simplicity, I limit attention to cases in which morality and self-interest require a particular action.

2 Nothing of importance to my argument turns on the questions that divide realists from antirealist or cognitivists from noncognitivists in ethics.

3 The overall verdict of ‘Reason’ could be complex. For example, it could be that several options are equally acceptable overall.

a worthwhile life.⁴ If morality were overriding, the demands of morality could not rationally be escaped. It may therefore seem liberating to recognize that moral reasons are not actually overriding.

I reject both of these points of view. As for the first, I will argue that morality is not discredited by its failure to override self-interest, for it is also the case that self-interest fails to override morality. As for the second, I will argue that there is no privileged standpoint from which to assess whether the demands of morality are ‘excessive.’ It is true that moral demands may appear excessive from the standpoint of self-interest, assuming, of course, that morality can demand actions contrary to one’s self-interest. This is not a telling point in my view, however, since self-interest does not override morality. Moral demands may appear excessive from other standpoints as well, standpoints that assess our lives as worthwhile, appealing, or excellent. I will argue, however, that these standpoints also do not override morality. It is never the case that we ought *simpliciter* to violate the demands of morality.

Before I can begin to explain and defend my view, I need to clear the ground of obstructions. I begin in section 2 by discussing my assumption that there can be conflict cases. In section 3, I distinguish two conceptions of self-interested reasons, a conception of self-grounded reasons and a conception of reasons to promote personal excellence. Section 4 defines the notion of overridingness in terms of normative importance. It also introduces the idea of ‘Reason-as-such,’ the idea of a standard for comparing the normative importance of morality and self-interest. Sections 5 and 6 argue that neither morality, nor self-grounded reason, nor a standpoint of personal excellence is to be identified with Reason-as-such. Section 7 argues that the idea of Reason-as-such is incoherent and that practical reason is disunified in a fundamental way. In section 8, I discuss whether this result discredits morality in any interesting way.

2. CONFLICT BETWEEN MORALITY AND SELF-INTEREST

Consider the myth of Gyges, as told by Plato in the *Republic*.⁵ Gyges is a shepherd in the service of the king of Lydia. One day he discovers a magic ring that makes him invisible when he twists it a certain way on his finger. Using the ring to become invisible at crucial times, Gyges commits

4 See Williams 1981, 22–24, 36–39. Also Slote 1983, 77–107; and Wolf 1982.

5 Plato 1992, 359d–360b.

adultery with the queen, attacks the king with her help, kills the king, and takes over the kingdom.

Plato's description of Gyges's situation is sketchy, but we can adjust the details as we like so that, for almost any moral theory and theory of self-interest, Gyges's situation was a case of conflict. To begin, imagine that Gyges did what he did because he believed that it would be in his self-interest. Now add details so that his belief was true. As a result of killing the king and winning the queen and the kingdom, he achieved a position in which he felt content with his life in a way that he had never felt before. With the queen, he enjoyed a fulfilling relationship of a kind he had never achieved before. As king, he developed talents of leadership and statecraft that he would not otherwise have developed. He valued his achievements. Given these details, which I hereby stipulate to be part of the story, I believe there is no room to deny that killing the king and winning the queen were in Gyges's self-interest, unless one accepts a highly implausible account of the nature of self-interest.

Next, we add details so that what Gyges did would count as morally wrong on almost anyone's account. Killing a person and violating a trust are at least *prima facie* wrong. Assume that the overall consequences of Gyges's actions were worse overall than the consequences would have been if he had remained a shepherd. Gyges was a good king, but not as good as the king he deposed and killed. The former king was benevolent, kind, and just. Gyges's intentions were morally indefensible; his plan made no room for the good of the king, the queen, or the kingdom, and considered only his own good. We can assume that Gyges realized he was in the wrong. Finally, Gyges's action expressed the vices of envy, greed, lust, and the desire for power. Again, I stipulate all of this, and so I believe there is no room to deny that Gyges's actions were morally wrong unless one has an implausible view about morality.

To be sure, some philosophers, including Plato, would argue that morality and self-interest must necessarily coincide. If these philosophers are correct, then, of course, unless Gyges's situation in the story is not possible, my assessment of it in terms of morality and self-interest must in some way be mistaken.

It might be argued, for example, that it is in one's self-interest to be as good as one can be, and that Gyges would have been a better person if he had remained a shepherd. I believe that arguments of this kind turn on an equivocation. It is in one's self-interest to achieve the best *for oneself*, but it does not follow that it is one's self-interest to be as *good* as one can be. It certainly does not follow that it is one's self-interest to be as *morally* good

as one can be. It would not have been better *for Gyges* if he had chosen to stay with his sheep.

Ethical egoism is the doctrine that a person is morally required to perform a given action just in case, of all the alternatives, the action would be most in the person's interest. Egoism implies that morality and self-interest cannot possibly conflict; it implies that Gyges did not do anything wrong, given that he acted in his self-interest. Egoism is quite implausible, however, and I assume it is false.

We surely do believe it is possible for morality and self-interest to diverge. Otherwise, for example, we would not bother to teach morality to our children as a system of norms distinct from self-interest.⁶ Of course, morality and self-interest may coincide in certain circumstances, but whether they do coincide is a contingent matter. It may be a fortunate truth that the pursuit of advantage never leads one morally astray. The Gyges example shows merely that this is not a necessary truth.

On my view, conflict situations involve conflict between normative verdicts. Some philosophers might deny this. They might claim that, to count as normative, a verdict would have to entail the existence of a reason, and they might deny that morality and self-interest are both sources of reasons. They might claim that there is a single basis for all reasons. On a Kantian approach, one might hold that reasons are considerations that would move fully autonomous agents. Moral considerations and considerations of self-interest are reasons only if they would move fully autonomous agents.⁷ On this view, conflict between reasons of these kinds would entail the existence of conflict in the motivations of fully autonomous agents, which one might argue to be impossible. On a Hobbesian approach, one might claim that all reasons are grounded in self-interest. One might infer from this that there are no reasons that conflict with reasons of self-interest.

It would take me too far afield were I to attempt to discuss these views in detail. There are, however, two things I would like to point out.

First, I do not need to insist that cases of conflict involve conflict between different kinds of *reasons*. All that I need to insist is that cases of conflict involve conflict between different overall verdicts as to what an agent *ought* to do, verdicts that are normative, or action-guiding. I merely assume the following: It is possible for the overall verdicts of morality and self-interest to conflict, where these verdicts are normative verdicts as to

6 This point is made by Scheffler (1992).

7 A view of this kind was suggested by Thomas Hill, Jr., in discussion.

what an agent ought to do.⁸ As I will explain, my understanding of the problem of overridingness depends on this assumption.

The second thing I need to point out is that, intuitively, morality and self-interest are both sources of reasons. We added details to Gyges's story to make it plausible that Gyges's killing the king was both morally wrong and in his self-interest. Given that Gyges's actions were in his self-interest, it is intuitively plausible that there were self-interested *reasons* for him to act as he did. And given that Gyges's actions were morally wrong, it is intuitively plausible that there were moral *reasons* for him not to do what he did. Gyges's situation illustrates the intuitive plausibility of the idea that there are moral reasons as well as reasons of self-interest and that there can be conflict between them.

3. REASONS, NORMATIVITY, AND SELF-INTEREST

The issue about overridingness is an issue about the relative normative significance of overall verdicts delivered by morality and self-interest. People can differ about what these overall verdicts would be in given cases, of course, and there are different theories about this. It is for this reason, in part, that the case of Gyges is so useful. We can agree that the case is a conflict case even if we do not agree in general about the notions of morality and self-interest.

Despite this, however, I cannot ignore the disagreement there is about the notion of self-interest. There are basically two views. These views agree that something is in a person's self-interest just in case it would be good for the person. They disagree about the content of the standards for assessing how well a person's life is going for her. A 'subjectivist' view might propose that a person's life is going well for her to the extent that her desires are being satisfied or her values are being fulfilled. More generally, a subjectivist view proposes a standard that evaluates a person's life on the basis of its relation to certain subjective psychological states of the person, such as the person's values, desires, or feelings of pleasure. An 'objectivist' view proposes certain other characteristics that contribute to how well a life is going, characteristics that are not matters of the relation between the life and the person's subjective states. For example, an objectivist view might judge a person's life on the basis of whether the person is developing

8 Strictly speaking, I may not need to assume the possibility of cases of conflict. For even if cases of conflict are not possible, there would still be the two kinds of verdict, and we could ask whether one kind overrides the other.

her talents. An objectivist view can be mixed, of course. For example, an objectivist view could give great weight in evaluating a person's life to whether the person is content with her life.⁹

It is possible to accept both subjective and objective standards without taking a position on the notion of self-interest. Indeed, standards of both kinds could be sources of reasons, although presumably reasons of different kinds.

In order to explain what I mean, I need to sketch an account of the nature of reasons. Nothing in my argument depends on the details of the account, but my formulation of the argument will make use of a central intuition about normativity.

The intuition is that if something has a normative property, such as the property of being right or wrong or good or bad, there are criteria that it meets or fails to meet; furthermore, a standard or a norm could in principle be formulated that calls for things of the relevant kind to meet the criteria at issue. For example, if an action is morally required, then it meets certain relevant moral criteria, and there is a moral standard or norm that calls on us to perform actions meeting these criteria. Of course, we can imagine various arbitrary standards, such as one calling on us to stand on our heads before eating any meal. Because of this, we need to add that the standards in question are not arbitrary. They are 'authoritative,' or 'justified.'

Elsewhere, I have proposed a model of the truth conditions of normative propositions that makes use of this idea.¹⁰ I assume that there are true propositions regarding what we are morally required to do and regarding what we ought to do in our self-interest. On my account, the truth of such a proposition depends on the existence of a relevant 'justified' or 'authoritative' standard.

I think there are reasons of many different kinds, including moral reasons, aesthetic reasons, and, presumably, reasons of self-interest. If there is a warranted, authoritative, or justified normative standard of a certain kind, and if this standard calls on people to choose in a certain way, then, I say, there are reasons of that kind for people to choose in that way.¹¹ If an agent is morally required to do something, my view implies that there is an authoritative moral standard that calls on her to do the thing. Given what I said about reasons, it follows that there is a moral reason

9 Richard Kraut discusses the idea of the good for a person (1994).

10 See Copp 1995.

11 This is an account of 'normative reasons.' For a similar view, see Smith 1994, 95.

for her to do it. Similarly, if an agent ought to do something because it would be in her self-interest, my view implies that there is an authoritative standard that calls on her to do things that are in her self-interest. It follows in turn that there is a self-interested reason for her to do the thing.

The source of the authority of standards of the various kinds is a substantive issue that is beyond the scope of this chapter. Given what I have said, it should be plain that a theory of moral judgment requires an account of the circumstances under which moral standards are relevantly authoritative.¹² Similarly, a theory of judgments of self-interest must provide a justification of the standard that calls on each of us to seek what is most in her interest; that is, it must provide an account of that in virtue of which the standard is relevantly authoritative.

Recall that I distinguished objectivist as well as subjectivist views about the content of standards of self-interest, standards as to how well a life is going for the person whose life it is. It is quite possible that some standards of both kinds are suitably justified, although presumably standards of the different kinds would be justified on different grounds. For simplicity, I will assume that this is so. Given the view I sketched about reasons, it follows that there are reasons of both kinds. I will assume, in particular, that each of us has reason to pursue a life in which his talents are fully developed and about which he feels content. I will call reasons of this kind ‘reasons of personal excellence.’ I will assume, moreover, that each of us has reason to pursue a life in which his desires or values are satisfied. I will call certain reasons of this kind ‘reasons of self-interest.’ Reasons of self-interest are a subclass of ‘self-grounded’ reasons. This is the next idea that I need to explain.

‘Self-grounded’ reasons are reasons grounded in a person’s own standpoint. I propose that the best life for a person, as assessed from her own standpoint, is the life that best meets her basic needs while also satisfying her values without frustrating her mere desires. A person’s basic needs, values, and desires together determine what reasons she has that are grounded in her own standpoint. I have developed this proposal elsewhere.¹³

12 Elsewhere, I have argued that moral standards are relevantly authoritative just when, roughly speaking, society needs their currency among its members in order to flourish. That is, their currency would be in the interest of the society as a whole. See Copp 1995.

13 See Copp 1995, ch. 9. See also chapter 10 in this book. The standard calling on me to satisfy my basic needs is not subjective in the sense I explained before. I cannot take the space to explain this here.

This is not exactly an account of self-interest. It allows that *any* of a person's values or desires can give her reasons, provided they do not conflict with her needs, even though some of her values and desires may not concern her own good. A person may have moral values, for example, and if so, my account implies that she has self-grounded reasons to realize these values. Hence, if Gyges had attached more value to keeping trust with the king, he might have had sufficient self-grounded reason to stay with his sheep rather than to try to take over the kingdom. This would not, however, have made it in his *self-interest* to stay with his sheep. Despite this, it seems to me that if a person has self-interested reasons grounded in the fact that she values or desires things for her own life, then her other values and desires must also ground reasons of the same kind. That is, a person's self-interested reasons are *among* the reasons she has that are grounded in her needs, values, and desires. I call this larger class of reasons *self-grounded*.

A person's values can be bizarre or self-destructive. It might be objected on this basis that the mere fact that a person values something gives her no reason to achieve it. Fortunately, nothing in my argument turns on the issue raised by this objection. Those who disagree with me will likely find that their view is addressed when I discuss reasons of excellence.

I assume, then, that there are both self-grounded reasons and reasons to pursue excellence. I submit that the question as to which of these kinds of reasons is more properly thought to be, or to include, 'self-interested' reasons is neither interesting nor substantive.

4. OVERRIDINGNESS

Let us return, then, to the questions about overridingness. What is a person to do when there is conflict between the action called for from the moral standpoint and the action called for from the self-grounded standpoint? Before we can begin to deal with these questions, we need to clarify the notion of overridingness. I interpret the notion as follows.

The claim that morality overrides self-grounded reason would be the claim that morality is *normatively more important* than self-grounded reason. We are assuming that Gyges morally ought not to have killed the king and yet that killing the king was best for Gyges in light of his needs, values, and desires. The claim that morality overrides self-grounded reason would imply that Gyges ought *simpliciter* to have done what he morally ought to have done.

To make sense of such a claim, we must suppose that there is a justified standard in terms of which to judge the relative normative significance of normative standpoints. This standard would specify criteria bearing on the normative importance of morality and self-grounded reason, or on the importance of their verdicts. The fact that morality is normatively more important than self-grounded reason, if it is a fact, would be the fact that morality meets the criteria specified by the standard in question, or that it meets the criteria more completely than does self-grounded reason.

This standard must of course be justified or authoritative. In addition, it must be *normatively more important* than any other standard that specifies criteria bearing on the comparative importance of morality and self-grounded reason. We want to know whether moral reasons override self-grounded reasons *period*, not merely whether moral reasons are overriding as assessed by some standard or other. We therefore need to know whether morality overrides self-grounded reason when assessed in terms of criteria specified by the normatively most important standard bearing on the comparison of standards. This standard must also bear on the choice of actions, for if morality overrides self-grounded reason when assessed in terms specified by this standard, then we ought *simpliciter* to act morally in every case of conflict.

If we ought *simpliciter* to act morally in every situation, then, on my view, there is a relevant authoritative standard that calls on us, in every situation, to act morally. I introduced this idea before. Now this standard could not simply be one among many. In order for its assessment of what we ought to do to be definitive and final, in order for it to settle what we ought to do *simpliciter*, it would have to be normatively the most important standard. We are not interested merely in whether, when a person is morally required to do something, she is required to do it by some standard or other.

In summary, then, the fact that morality overrides self-grounded reason, if it is a fact, would consist in the fact that morality is *normatively more important* than self-grounded reason as assessed in terms of the *normatively most important* standard bearing on the comparison of standards and the choice of actions. We could call this standpoint or standard, if it exists, *Reason*, or *Reason-as-such*. Then we could say that morality overrides self-grounded reason just in case it is never Rational (knowingly) to do otherwise than act morally in situations of conflict.¹⁴

14 Scheffler says that the “claim of overridingness” is “the claim that it can never be rational knowingly to do what morality forbids” (1992, 52).

The standard of Reason-as-such would have the following properties, which, for future reference, I will call ‘comprehensiveness’ and ‘supremacy’: First, there are various special standpoints or standards for choice – standpoints such as that of morality, self-interest, prudence, etiquette, law, or aesthetics. The standard of Reason would take the verdicts given by all the special standpoints regarding any situation where an agent needs to choose; it would evaluate these verdicts without any question-begging; and it would produce an overall verdict as to what the agent is to do. As I will say, it would be ‘comprehensive.’ Second, the standard of Reason would be the normatively most important standard for assessing such verdicts and the choice of how to act. Hence, an agent *ought simpliciter* to comply with its overall verdict. Reason-as-such would not be merely another standpoint alongside the special standpoints. As I will say, it would be ‘supreme.’

The issue whether there is such a thing as Reason-as-such is not the same as the issue whether morality overrides self-grounded reason. For Reason-as-such might fail to align necessarily with either morality or self-grounded reason. It could be that although it sometimes aligns with morality, and sometimes aligns with self-grounded reason, it sometimes aligns with neither. If I am correct, however, the existence of Reason-as-such is a necessary condition of its being the case that morality overrides self-grounded reason. For if morality overrides self-grounded reason, there is a comprehensive and supreme standard such that, necessarily, in every case of conflict, it yields the verdict that each agent is to comply with the overall verdict of morality. It would follow that each agent *ought simpliciter* to do the morally right thing.

5. MORALITY, SELF-GROUNDED REASON, AND REASON-AS-SUCH

Assuming that there is a standard of Reason-as-such, there are three possibilities. First, it may be that the verdicts of morality are verdicts of Reason-as-such, and the standpoint of Reason includes morality; second, it may be that self-grounded verdicts are verdicts of Reason-as-such, and Reason includes the self-grounded standpoint. On either of these views, Reason-as-such is seen as simply a wider standpoint than the moral standpoint or the self-grounded standpoint, respectively. The third possibility is that the standpoint of Reason is different in character from both of these other standpoints. This possibility leaves open whether Reason necessarily agrees with morality or with self-grounded reason.

There is something to be said for each of the first two possibilities. Consider the idea that the moral standpoint is the supreme standpoint in cases where morality renders a verdict. One might claim it is analytic, or conceptually guaranteed, that if morality requires Gyges not to kill the king, all things considered, then Gyges ought *simpliciter* not to kill the king, and there is no further question as to what he ought to do. One might add that it is analytic, or conceptually guaranteed, that if morality requires Gyges not to kill the king all things considered, then since Gyges ought *simpliciter* not to kill the king, he *would* refrain from the killing if he were *fully rational*. On this view, in cases where morality yields an overall verdict, its verdict is identical with that of Reason-as-such. Call this view ‘moral rationalism.’¹⁵

Consider now the idea that the self-grounded standpoint is the supreme standpoint, the standpoint of Reason-as-such, in cases where self-grounded reason renders a verdict. One might claim that it is analytic, or conceptually guaranteed, that if Gyges’s needs, values, and desires argue that it was best for Gyges to kill the king, then it was *rational* of Gyges to do so, and there is no further question. He simply ought to kill the king. On this view, in cases where the standpoint of self-grounded reason yields an overall verdict, it is identical with that of Reason-as-such. Call this view ‘self-grounded rationalism.’

The reply to these suggestions is essentially that although they purport to answer the question about overridingness, in fact they do not. For, in each case, the question could be reformulated. Let me explain.

Moral rationalism contends that it is analytic that if one morally ought to do something all things considered, then one ought to do the thing *simpliciter*, and there is no further question as to what one ought to do. I agree, of course, that when a person ought morally to do something, we often speak simply of what the person ‘ought’ to do, without mentioning the qualification that this is what she ought *morally* to do. But the fact that this qualification is not mentioned hardly settles what an agent ought to do in a case of conflict between morality and self-grounded reason. We can ask, What is the thing that the person ought to do from a standpoint that provides a definitive assessment of the relative normative priority of morality and self-grounded reason? Or is there no such standpoint?

Moral rationalism also claims that it is analytic or conceptually guaranteed that if one morally ought to do something, all things considered,

15 Smith defends a similar view (1994, 85–91, 130–202, esp. 182–184).

then one will do it if one is ‘fully rational.’ But in the absence of an argument that the ‘fully rational’ person in this context is the person who does exactly what is required by Reason-as-such, this claim merely side-steps our question. Our question could be reformulated as follows: Consider a case where there is conflict between what a person would do if she were ‘fully rational’ (i.e., what she morally ought to do) and what she ought to do from the self-grounded standpoint. What is the thing she ought to do *simpliciter*, from a standpoint that provides a definitive assessment of the relative normative priority of morality and self-grounded reason?

It might be replied that no requirement counts as a *moral* requirement unless it is overriding. On this view, of course, there is no problem of whether morality is overriding. However, for any *putative* moral requirement, or ‘schmoral’ requirement, there is a problem of whether it is in fact a moral requirement. On this view, there may be no moral requirements since schmoral requirements may not be overriding. The question we are interested in could be expressed as follows: In a case where there is conflict between what a person ought ‘schmorally’ to do and what she ought to do in light of her needs, values, and desires, is her ‘schmoral’ requirement a moral requirement? What is the thing she ought to do from a standpoint that provides a definitive assessment of the relative normative priority of these two verdicts?

Self-grounded rationalism claims that the thing that a person ought to do in light of her needs, values, and desires is the ‘rational’ thing to do. I agree, of course, that this is a common way of speaking. We do at least frequently express verdicts as to what would be in our self-interest as verdicts about what it would be ‘rational’ to do. This does not settle the question about overridingness, however, for the question could be reformulated: Consider a case where there is conflict between what a person morally ought to do and what she would be ‘rational’ to do, given her needs, values, and desires. What is the thing she ought to do from a standpoint that provides a definitive assessment of the relative normative priority of these two verdicts?

There is a substantive question we want to ask, and the views I have been considering do not answer it. They may appear to answer it, but in each case the question arises again, even if in different terminology.

Here is a way to think of the problem. Morality consists of a particular system of justified standards. These standards call for certain things; they call for Gyges not to kill the king. Similarly, the norm of self-grounded reason calls for agents to pursue the satisfaction of their needs, values, and desires. This norm calls for Gyges to kill the king. The question

about overridingness is about the relative normative importance of these two systems of norms. It cannot be answered on the basis of either system. Each of these systems is concerned to evaluate our actions, traits of character, and the like; neither is concerned with the issue about the relative significance of normative standpoints. Moreover, although each system can yield a verdict about verdicts of the other system, these verdicts do not settle the relative normative significance of the systems. Morality would prescribe that we comply with our moral duty rather than act on our self-grounded reasons in cases of conflict, but self-grounded reason would prescribe that we act on our self-grounded reasons in such cases. These verdicts leave unanswered the key question as to which of these verdicts is normatively the more important.

I conclude, then, that in order for morality to override self-interest, or vice versa, there would have to be some other normative standpoint that ranked one as normatively more significant than the other. And this standpoint would have to be normatively more significant than either morality or self-grounded reason; it could not be merely another standard for choice alongside morality and self-grounded reason that gives rise to merely another special kind of reasons. This supreme and comprehensive standard, if it exists, would be the standard of ‘Reason,’ or ‘Reason-as-such.’

I believe there is no such thing. But before I attempt to show this, let me explore the credentials of the standpoint of personal ‘excellence.’ Some philosophers would identify this standpoint with self-interest. I want to consider whether it can be identified with Reason-as-such.

6. THE STANDPOINT OF PERSONAL EXCELLENCE

Several recent discussions of the issue whether morality is overriding, and of the related issue whether morality is ‘too demanding,’ have invoked ideas about the admirable, desirable, or excellent life. Bernard Williams discusses the importance of our “projects” to the “rational justification” of our choices. In discussing a fictionalized Gauguin, who wrongfully abandoned his family in order to go to the south Pacific to pursue his career as a painter, Williams appears to say that Gauguin made the best choice.¹⁶ Williams’s thought may be that morality asked too much of Gauguin, for Gauguin could not reasonably have been expected to sacrifice success in

16 Williams 1981, 22–24, 36–39. Williams’s discussion is difficult to follow, so I cannot claim to have the only or the best interpretation.

the central project of his life in order to take care of his family. Michael Slote argues that Gauguin receives our *admiration* even though the central choice of his life was morally wrong.¹⁷ Perhaps Gauguin deserved our admiration because his choice to pursue his painting made his life more fulfilling and successful than it would otherwise have been. Susan Wolf discusses the desirability of living an *interesting* and *appealing* life, and she argues that a morally flawless life would not generally be an especially appealing one to others.¹⁸ She might therefore say that even though Gauguin's life was *morally* flawed as a result of his decision to abandon his family, he lived a more appealing or interesting life than he would have if he had decided otherwise. Similar things could perhaps be said about Gyges, for he too had a 'project' that morality asked him to abandon, and he lived a more interesting life than he would have, if he had complied with his moral duty. One might admire him for his success in his project.

Williams, Slote, and Wolf appear to be invoking a standard that is distinct from morality, a standard concerned with the choice of the best life – the life that would be best *for* the person whose life it is. Let me simplify by assuming that they have the same fundamental idea in mind, a standard of 'personal excellence.' Each of them appears to be suggesting that morality deserves equivocal support at best from this standpoint of personal excellence.¹⁹

An Aristotelian view of personal excellences is a familiar one, and I will suppose, for the sake of argument, that an Aristotelian view is at work in the thinking of Williams, Slote, and Wolf. In particular, I will assume that the following standard is at work: We are to pursue projects that, if pursued, would develop our most valuable talents as fully as they could be developed while also giving us enjoyment. The best life for a person is a life of success in enjoyable self-developing projects.

Given this framework, we see immediately that the standard of excellence is distinct both from the familiar standards of morality and from the standard of self-grounded reason. Moreover, it is arguable that the pursuit of excellence does not necessarily align either with morality or with self-grounded reason.

17 Michael Slote discusses "admirable immorality" (1983, 77–107).

18 See Wolf 1982.

19 The idea of the good for a person is discussed by Kraut (1994). Hurka has developed an account of personal excellences (1994).

It is plausible, for example, that although Gauguin's choice was morally wrong, it was recommended by the standard of excellence. Painting was the central project of Gauguin's life, and his talent for painting was arguably the most valuable of his talents. As for Gyges, I stipulated that his action of taking over the kingdom enabled him to develop talents of leadership and statecraft that he would not have been able to develop as a shepherd. It is arguable that these talents are more valuable than the talents a shepherd could develop, and, in any case, Gyges had already developed the latter talents. Arguably, then, the standard of excellence supported the actions of both Gauguin and Gyges despite their being morally wrong. Apparently, then, excellence does not necessarily align with morality.

Self-grounded reason calls on a person to choose what would be best for her from her own standpoint, in light of her needs, values, and desires. The standard of excellence, however, calls on a person to choose what would most facilitate a life of enjoying self-developing projects. Such a life is not necessarily best from the person's own standpoint, given her own values. The projects she values may not be projects that would best develop her talents. If Gyges had most valued his work as a shepherd, then continuing to work as a shepherd would have been recommended by self-grounded reason even if the pursuit of political power would have been recommended by the standard of excellence. It is certainly arguable, then, that the best choice from the standpoint of self-grounded reason is not necessarily the same as the best choice from the standpoint of personal excellence.

These claims about personal excellence would of course be debated by some philosophers. Some would argue that a person living the most excellent life would necessarily be living a morally virtuous life. For our purposes, however, the crucial issue is whether the standard of excellence can be identified with that of Reason-as-such. Does this standard have the two properties of comprehensiveness and supremacy?

The standpoint of excellence does yield verdicts about at least some of the verdicts of morality and self-interest. It may therefore be comprehensive. The key question, however, is whether it is supreme. Is the standpoint of excellence normatively the most important standpoint? Unless this is so, it is not the case, when a person is required to do something by the standard of excellence, that she ought to do it *simpliciter* and without qualification. And unless this latter thing is so, then the fact, if it is a fact, that the standard of excellence aligned in Gyges's case with self-grounded reason would have no tendency to show that morality is overridden in this

case by self-grounded reason. It would have no tendency to show that Gyges ought *simpliciter* to kill the king.

Intuitively, the standpoint of excellence is merely a special standpoint of evaluation. What we ought to do in order to best develop our talents is not what we simply ought to do, without any qualification. If this is correct, the standpoint of excellence is not supreme among all special standpoints in the way that Reason-as-such would be, if it existed.

The claim that the standpoint of excellence has the property of supremacy is the claim that it is supreme over other standpoints as assessed in terms of certain relevant criteria. Let these criteria be specified by a standard S. The first point I want to establish is that the standard of excellence is not itself this standard S. The standard of excellence speaks to the relative excellence of our choices, not to the relative normative importance of various standards for choice. In some cases it does yield verdicts about the verdicts of other standards; it prescribes that we pursue excellence in cases of conflict, rather than that we comply with either our moral duty or self-grounded reason. But morality also delivers a verdict in such cases; it prescribes that we comply with our moral duty rather than pursue excellence. The verdict of the standard of excellence in favor of the pursuit of excellence leaves unanswered the key question of whether this verdict is normatively superior to the (perhaps) contrary verdict of morality. This question cannot be answered by the standard of excellence. To suppose otherwise would be to argue in a circle. Hence, the standard of excellence is not standard S.

It follows that the standard of excellence does not have the property of supremacy. For in order to yield definitive assessments of the relative importance of normative standpoints, S must be supreme. It must take the verdicts of the various standpoints and yield verdicts about them, and these verdicts must be definitive in the way that only the verdicts of the normatively most important standpoint could be. Hence, the claim that the standard of excellence is supreme has led to the conclusion that it is not supreme, that S is supreme. It follows that we must deny the claim and conclude that the standard of excellence is not the standard of Reason.

7. SKEPTICISM ABOUT THE UNITY OF REASON

When we deliberate about what to do, we may try to take into account all relevant considerations and to make the best or the right decision. We may not want to make a decision that is merely right or best from one

standpoint. We may want to make the decision that is best *period*.²⁰ This would be the decision required by Reason-as-such.

To be sure, a morally virtuous person wants to do what would be morally best, but she then presumably thinks that to do what is morally best *is* to do what is best, period. At least, she does not think that something else would be best, period. This is part of what it is for her to be morally virtuous. Similarly, a self-interested person wants to do what would be best for herself, from her own standpoint. But she presumably thinks that what would be best for her from her own standpoint *is* the best thing for her to do. At least, she does not think that something else would be the best thing for her to do, period. Otherwise, she would not be wholly self-interested, for she might think she could do better than pursue her self-interest.

I have claimed that there is no standard of Reason, no supreme and comprehensive standpoint. If we accepted this position, we would have to give up the thought that when we deliberate, we can make the best or right decision, period. For this thought presupposes that there is a standard for the evaluation of choices that is normatively the most important. This would be the standard I have been calling the standard of Reason or Reason-as-such, the standard that yields verdicts as to what we ought to do *simpliciter*, and without any qualification.

To be sure, I have not *shown* that there is no such supreme and comprehensive standard; for all I have argued, then, it is possible that one does exist. But I have argued that morality is not identical with Reason, nor is self-grounded reason, nor is the standard of excellence. These standards are all warranted in their own ways, and they give rise to reasons of their respective kinds, or so I assume, but they are not normatively supreme.

I believe that there is no such thing as Reason-as-such. First, I do not believe we have any clear conception of what such a thing would be. Consider any candidate for the standard of Reason. Call it S. It is quite unclear what status S could have that would give it the kind of supremacy it would need in order to qualify as the standard of Reason. It is also unclear what status it could have that would give it authority, such that it would be indeed a source of reasons, without giving it a special perspective on the facts of situations where choices must be made. We would then see S as simply another special standard rather than as the one standard

20 Philip Clark stressed this point, in discussion.

qualified to appraise the significance of the verdicts of all of the special standpoints.

Second, there is the following *reductio* of the idea that our candidate S has the property of supremacy. The claim that a standpoint has the property of supremacy is the claim that it is the *normatively most important* standpoint. I argued before that comparisons of the relative normative importance of standpoints must be made in terms of criteria specified by some relevant authoritative normative standpoint. Hence, the claim that the candidate S has the property of supremacy is the claim that it is normatively more important than any other standpoint, as assessed from a relevant authoritative standpoint. That is, if S is normatively the most important, then there is some authoritative standard R that yields the verdict that S is normatively the most important standpoint. R assesses the relative significance of S's verdicts and the verdicts of all the special standpoints, and R determines that the verdicts of S are definitive as to what we ought to do *simpliciter*. Now, either standard R is identical to S or it is not.

We cannot suppose that R is identical to S. For a standard cannot be normatively the most important in virtue of its meeting criteria it *itself* specifies as criteria to be met by standards. The standard of Reason is to be normatively the most important *simpliciter*, not merely normatively the most important from its own standpoint. Morality prescribes that we ought to do our moral duty, and it judges that standards that prescribe otherwise are morally wanting by comparison with itself. But this fact about morality does not suffice to make it the case that morality is the normatively most important standpoint *simpliciter*; at best it shows that morality is the *morally* most respectable standpoint. Similarly, the standard of self-grounded reason is presumably the most respectable from the standpoint of self-grounded reason. And for any candidate S, there is presumably a sense in which S assesses the significance of its own verdicts by comparison with the verdicts of other standpoints; for it prescribes, I assume, that we ought to perform exactly the actions it prescribes. But this does not suffice to make it the case that S is normatively the most important standpoint *simpliciter*; it merely shows that S is the most acceptable standpoint from its own standpoint. This verdict of standard S leaves unanswered the key question as to whether the verdict of S is normatively more important than all other verdicts. To suppose otherwise would be to argue in a circle. Therefore, if S is normatively the most important standpoint, then it meets criteria specified by some authoritative standpoint R that is distinct from S. This is what constitutes S as normatively the most important.

This standard R must be normatively the most important standard. Otherwise its verdict would not settle definitively the relative normative status of S and the special standpoints. Otherwise, there would be some standpoint superior to R, and *its* assessment of the relative importance of S and the special standpoints would be the definitive one. But then, if R is normatively more important than S, it follows that S is not in fact the normatively most important standpoint.²¹ Moreover, although we ought *simpliciter* to do what S prescribes, this is a judgment made from the standpoint of R, not a judgment made from the standpoint of S itself. Hence, S itself is not to be identified with the standard of Reason-as-such, which contradicts the assumption with which we began.

The same *reductio* can be run on the assumption that R is the standard of Reason-as-such. It appears, then, that the idea of a standard of Reason-as-such is incoherent.

The incoherence can be displayed in two sentences: The claim that a standard S has the property of supremacy is the claim that it is normatively the most important standard as assessed in terms of some other standard, R, which is the normatively most important standard. But only one standard could be normatively the most important.

When we are trying to decide how to act, we do sometimes manifestly have the thought that there is something that it would be best to choose, period. This thought commits us to the existence of a standard that determines the proper weight of all the reasons that bear on a decision and that therefore determines what would be best or right *simpliciter*. If I

21 Let me briefly discuss two possible objections to the conclusion that R must be normatively more important than S. (1) Perhaps R is exactly as important normatively as S. But then, what would constitute them as exactly equally important? It could not be the assessment of R by S and vice versa. This is simply to argue in a larger circle than we argue in if we suppose that R is identical to S. There would then have to be some additional standard T that specifies criteria such that R and S are equally important. Now there is the problem of the relative status of R, S, and T, which raises the very issues that were supposed to be escaped by supposing that R and S are equally important. (2) Perhaps R and S are normatively 'incomparable.' But then all we have is that a standard of indeterminate significance ranks S as superior to all other standpoints. This is not sufficient to constitute S as normatively superior *simpliciter*; at most it shows that S is superior from the standpoint of R. Moreover, there might be some other standpoint T that is superior to R and that assesses S as normatively inferior to some other standpoint. To eliminate this possibility, we would have to suppose that no standpoint is superior to either R or S. But then there must be some other standard that specifies criteria of evaluation according to which no standard is superior to R or to S. This begins the regress anew. I see no way to avoid a regress problem.

am correct, this thought is false.²² There are only the various reasons of the various special kinds, and in weighing them we are simply deciding which to act on. Our decision may be *guided* by the reasons there are, but it is not *determined* by the reasons. That is, again, the reasons do not balance out from a standpoint that determines their correct weight and the right choice, period and without qualification.

My position is a skepticism about the unity of practical reason – “skepticism” seems the appropriate term, for although I deny that practical reason is unified, I concede that I have not proven this. My argument depends on the assumption that there are various special kinds of normative considerations, including those of morality and self-grounded reason, and that it is possible for such considerations to conflict. It also depends on the intuitions about normativity and reasons that I introduced earlier in the chapter.

Henry Sidgwick also doubted the unity of practical reason. He thought it unlikely “that the performance of duty will be adequately rewarded and its violation punished.” In consequence, he said, there is a “vital need that our Practical Reason feels of proving or postulating this connexion of Virtue and self-interest, if it is to be made consistent with itself. For the negation of the connexion must force us to admit . . . that the apparently intuitive operation of the Practical Reason . . . is after all illusory.”²³

It may seem that matters are different when it comes to the choice of what to believe. It may seem that reasons of the kind that ground our knowledge, which we may call ‘epistemic’ reasons, simply do override all other reasons that bear on what to believe. I doubt that this is correct. From the ‘epistemic’ standpoint, we have reason to believe a proposition if it is more likely to be true than its negation, given the available information.

22 Hence, in *most* situations of deliberation it is false that there is something we could choose that would be best, period. There may be *some* situations in which this thought is true, however, for if all the special reasons speak in favor of the same option, then any plausible candidate for the standard of Reason would select that option as best, period. Compare the problem of social choice that was explored by Kenneth Arrow (1963). Arrow argued there is no function from individual preference rankings over social options into a social preference ranking over those options – no function that meets certain theoretically desirable conditions. I am arguing that there is no standard that takes the rankings over sets of options given by the various special standards for choice into a supreme ranking over these sets of options – no standard that meets certain theoretically desirable conditions, including especially that it be comprehensive and supreme. In this note, I am pointing out that in certain special circumstances only one ranking could plausibly be accepted. In particular, if all the special standards rank the options the same way, then that is the only way that the supreme standard or Reason could rank the options.

23 Sidgwick 1907, 496–509. The quoted passages are on pp. 508 and 509.

But there are other kinds of reasons to believe things. In some cases, for instance, there may be self-grounded reasons to believe something, and these reasons may conflict with the epistemic reasons. Suppose, for example, that a person has certain values about the kind of person she wants to be. She values being a ‘God-fearing’ person with traditional moral standards. Or she values having faith that God exists. In either case, her values give her reason to believe that God exists; they give her a self-grounded reason to believe that God exists. Reasons of this kind may conflict with epistemic reasons.²⁴ It is not obvious in such cases what a person ought *simpliciter* to believe even though it may be obvious what she ought *epistemically* to believe.

Epistemic reasons override self-grounded reasons only if there is some third standard for deciding how to believe, a standard that is comprehensive and supreme. It is not obvious that there is such a standard. We should not restrict our skepticism about the unity of reason to so-called ‘practical reason.’

To be sure, conflicts between kinds of reasons to believe are not generally a problem. To believe something is to accept it as true. Because of this, the epistemic reasons for believing a proposition are central to the question whether to believe it, for they speak to the issue whether the proposition is more likely to be true or false. Normally a person simply *does* believe a proposition if she believes it is most likely to be true given the available information; for if she believes this about the proposition, it is a small step to accept the proposition as true. It *is* a step, however, and it is not necessarily the case that a person takes the step.

Notice that it is not paradoxical for a person to believe a proposition even though she concedes its negation is more likely to be true given the available information. She may suspect that the evidence is misleading,

24 Some philosophers would object that, no matter what proposition *p* is at issue, the only reasons to believe that *p* are epistemic reasons, and that practical reasons of the kind I discuss in the text are reasons to attempt to get oneself to believe (or to continue to believe) that *p* rather than reasons to believe that *p*. I am grateful to George Nakhnikian for helpful discussion of this issue in personal correspondence. To take account of this objection, I would need to reformulate the issue I am addressing in the text, but the issue would still arise. A person might see that it would be disloyal of her to believe that *p*, yet she might also see that there are good epistemic reasons to believe that *p*. Out of loyalty, she might ‘refuse’ to believe that *p*. Suppose she could take a pill that would cause her to believe whatever she most wanted to believe at the moment she took the pill. In this case, she could cause herself to believe that not-*p*. In order to decide ‘what to believe,’ she would have to weigh her epistemic reasons to believe that *p* against her practical reasons to make herself believe that not-*p*. These reasons ‘conflict’ in an obvious way. [This note was added in 2006.]

for example. Or she may be moved by some nonepistemic reason, such as a self-grounded reason. She may believe that God exists because she sees this belief as an expression of the virtue of faith even though she concedes that it is not likely that God exists given the evidence. She may see the absence of evidence as a test of her faith.

Therefore, the plurality of kinds of reasons that have a bearing on our decisions as to how to act is mirrored in a plurality of kinds of reasons for belief that have a bearing on our decisions as to what to believe. There is no supreme comprehensive standard of Reason-as-such in either case – or, at least, I believe there is no such standard.

8. THE STANDING OF MORALITY AND OF SELF-INTEREST

What does the denial that reason is unified imply about the relative standing of morality and self-interest? If neither morality nor self-interest overrides the other, does it follow, for example, that there is no reason to be moral? Does it follow that there is no answer to the question, “Why be moral?”

To begin with, my view does not challenge either the facts as to what we ought morally to do or their normativity, their bearing on our choices of how to act. Indeed, the question whether morality overrides self-interest presupposes that morality yields verdicts as to what we ought morally to do. The correct moral verdicts, whatever they are, have a bearing on our choices of how to act because they are verdicts as to what we *ought* to do. They imply that there are reasons to act as we morally ought.

Moral reasons can of course be given for choosing to do the morally required thing. In cases of conflict, however, self-interested reasons can be given for choosing *not* to do the morally required thing, and I believe there is no well-grounded answer to the question, What ought we to do *simpliciter*? If this is correct, we might ask ourselves, Why should we be moral?

This question is normally understood as the question whether we have sufficient self-interested reason to be moral. It may seem that if morality were overriding, morality would not need the support of self-interest, and the question “Why be moral?” would not be interesting. But, on the contrary, even if morality did override self-interest, it would still be important to determine whether there are good self-interested reasons to be moral simply because people are typically moved to act in their

self-interest. The question “Why be moral?” has an importance that is independent of the issue of overridingness.

It is part of my view that morality and self-interest are in similar positions. The question “Why act as self-interest requires?” is just as pressing, theoretically, as the question “Why be moral?” The “Why be moral?” question is the more pressing given the facts of human psychology, however, and the important standpoint from which to answer it is the standpoint I have called ‘self-grounded’ reason, as I will explain.

A person normally does the thing that would be recommended by self-grounded reason, given the agent’s own perception of her needs, values, or desires. This is a thing she perceives she needs to do, or it is a thing that she thinks would best promote satisfaction of her values or desires. Hence, typically, although not necessarily, when a person does what she is morally required to do, her action is recommended by self-grounded reason. Indeed, a person who acts morally acts on her moral values. That is, she subscribes to the moral standards that correspond to the overall moral verdict as to what she ought to do, and this explains her action. In such cases, a person may in fact be acting in the way that is best supported by her self-grounded reasons. I earlier explained that, in my view, reasons of self-interest are a subset of self-grounded reasons. A person who acts morally may be acting on the balance of *self-grounded* reasons, even if her action is not in her self-interest, narrowly construed.

Given all of this, and given our interest in having people act morally, it is in our interest that people subscribe to moral standards and, in particular, that they subscribe to justified moral standards. Moral education and the informal moral sanctions that reinforce our moral values are therefore of central importance to us all. For if people have the proper moral values, if they subscribe to justified moral standards, they thereby have self-grounded reasons to act morally. And from a person’s own perspective, self-grounded reasons are as compelling, and they are compelling on the same basis, as self-interested reasons narrowly conceived.

If, therefore, we can create social conditions in which people have the proper moral values, we can thereby contribute to bringing about a situation in which there is less conflict between morality and self-interest than would otherwise be the case, and in which the remaining conflict is less important. For we can thereby help bring it about that people have self-grounded reason to act morally. Of course, if there is conflict between people’s ability to meet their basic needs and the demands of morality, then the balance of self-grounded reasons may still speak against

morality. But this is no reason to despair of the place of morality in rational decision-making. It is rather a reason to favor creating a society in which people are enabled to meet their basic needs. A society in which people are enabled to meet their needs while also being encouraged to have proper or justified moral values would be one in which people on the whole would have self-grounded reason to act morally. In such a society, morality and self-interest would walk the same road.

REFERENCES

- Arrow, Kenneth. 1963. *Social Choice and Individual Values*, 2nd ed. New Haven, Conn.: Yale University Press.
- Copp, David. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- Hurka, Thomas. 1994. *Perfectionism*. New York: Oxford University Press.
- Kraut, Richard. 1994. "Desire and the Human Good." *Proceedings and Addresses of the American Philosophical Association* 68, no. 2: 39–54.
- Plato. 1992. *Republic*. Trans. G. M. A. Grube and C. D. C. Reeve. Indianapolis, Ind.: Hackett.
- Scheffler, Samuel. 1992. *Human Morality*. New York: Oxford University Press.
- Sidgwick, Henry. 1907. *The Methods of Ethics*, 7th ed. Rpt., New York: Dover, 1966.
- Slote, Michael. 1983. *Goods and Virtues*. Oxford: Oxford University Press, 1983.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Williams, Bernard. 1981. "Moral Luck." In *Moral Luck*, 20–39. Cambridge: Cambridge University Press.
- Wolf, Susan. 1982. "Moral Saints." *Journal of Philosophy* 79: 419–439.

The Normativity of Self-Grounded Reason

1. INTRODUCTION

In this chapter, I propose, and then attempt to ground, a standard of practical rationality. According to this standard, to a first approximation, rationality consists in the efficient pursuit of what one values.¹ This standard differs from the familiar principle of instrumental reason, which requires us to take the most efficient means to our ends, for it gives special emphasis to those of our ends that qualify as our values. It also differs from the principle of self-interest, which requires us to pursue our own good, both because we might value the good of others as much as our

I am grateful to many people for helpful discussion over the many years in which I have been developing the ideas in this chapter. With apologies to those whose help escapes my memory, I would like to thank Nomi Arpaly, Sam Black, Michael Bratman, Justin D'Arms, Dan Farrell, Pat Greenspan, Don Hubin, Dan Jacobson, Marina Oshana, Michael Ridge, Michael Robins, David Sobel, Pekka Väyrynen, and David Velleman. I presented early versions of some of the ideas in this chapter to audiences in the departments of philosophy at the University of Alberta, the University of Maryland at College Park, l'Université de Montréal, the University of Southern California, and the University of Florida, at the 1999 Conference on Moral Theory and Its Applications, Le Lavandou, France, and at the 2001 Conference on Reason and Deliberation, Bowling Green State University. The current version of the chapter was presented to the 2004 Conference on Personal Identity, organized by the Social Philosophy and Policy Center, Bowling Green State University. Michael Smith and the students in his 2005 graduate seminar at Princeton University discussed this chapter with me, and I hope to be able to take their helpful criticisms and suggestions into account in future work. I am grateful for the helpful comments of those who participated in the discussions on all of these occasions and especially to the other contributors to the volume of *Social Philosophy and Policy* in which this chapter was originally published, and its editors. I owe special thanks to Ellen Paul for encouraging me to integrate my thinking on identity with my thinking on rationality and for her useful comments.

1 This formulation ignores a qualification I have discussed in Copp 1995, ch. 9, and Copp 2001.

own good, and because, if we are unwise, we might value things that are bad for ourselves. I speak of the conception of rationality I develop as “self-grounded” because it requires the pursuit of a person’s *own* values, and also because, as I shall argue, a person’s values are grounded in her *identity*, on one useful conception of the identity of persons.

The idea of *grounding* a standard of rationality will require some discussion. The term comes from Immanuel Kant, as does my strategy, broadly understood, for I aim to ground the standard of self-grounded reason in the idea of autonomous agency, which of course is a strategy inspired by Kant’s work in the *Groundwork of the Metaphysics of Morals*.² The account I shall give of autonomous agency is very different from Kant’s, however, and my conception of rationality is modest by comparison with a Kantian conception.³ Most importantly, it does not guarantee a strong Kantian link between rationality and morality, for a person can exhibit a rational efficiency in the pursuit of what she values even if her values are morally abhorrent. Moreover, the grounding that I shall propose is compatible with naturalism; it does not rest on claims that purport to be synthetic a priori, nor does it depend on a nonnaturalistic metaphysics.

The basic idea is this: To be autonomous is to be *self-governing*. To be rational is at least in part to be *self-governing*; it is to do well, by a standard that we need to specify, in *governing oneself*. I argue that a person’s values are aspects of her *identity* in a way that most of her ends are not, and that it therefore is plausible to view action governed by one’s values as *self-governed*. This is also plausible on independent grounds. Given this, I say, rational agents comply with a standard – the ‘values standard’ – that requires them to serve their values, and to seek what they need in order to continue to be able to serve their values.⁴ I argue, then, that there is *reason* for an agent to serve her values and to seek what she needs in order to continue to be able to serve her values. An agent’s values are a source of reasons – reasons of a kind that any rational person would take into account in deliberation, if she were aware of them, just in virtue of

2 Kant 1785, 50 (Ak 448). In references to this work, I refer first to the pagination of the Ellington translation and then, in parentheses, to the pagination of the Prussian Academy Edition (Berlin: de Gruyter, 1902–).

3 Christine Korsgaard has recently suggested that “the normativity” of the principle of instrumental rationality “must be traced to the agent’s self-government, specifically to his capacity to be motivated to shape his character in accordance with an ideal of virtue” (Korsgaard 1997, 220 n. 13). My account does not invoke a moralized conception of autonomy.

4 In speaking of “serving” a value, I mean to cover both cases of taking means that are instrumental to fulfilling or realizing something one values and cases of doing what is constitutive of fulfilling or realizing a value. For this usage, see Lawrence 2002, 90.

being rational. I call such reasons “self-grounded reasons,” or “reasons of autonomy.” I sometimes call the conception of self-grounded rationality the “autonomy conception.”

The autonomy conception combines an account of the *content* of the standard of rationality with an account of its *grounding*. I begin, in section 2, by outlining some of the advantages of the conception. In sections 3 and 4, I explain the conception of autonomous agency, and I briefly discuss Michael Bratman’s model of such agency. In section 5, I develop an account of the identity of persons, and in section 6, I propose a conception of an agent’s values and link it to my account of identity and through it to the conception of autonomous agency. I then turn to the content of the standard of rationality. In section 7, I outline the values standard, and, in section 8, I address an important objection, the objection from ungrounded ends. In sections 9 and 10, I explain the idea of grounding a standard and argue that the values standard can be grounded in the conception of autonomous agency. Finally, in section 11, I explain the way in which, on the autonomy conception, rationality has priority for practical deliberation.

2. WHY TAKE THIS ROAD?

I believe that the autonomy conception has both intuitive and theoretical advantages over the familiar alternatives. One advantage is that it is compatible with a pluralistic view of reasons and normativity, a view that could be called “reasons pluralism.” I do not aim to defend reasons pluralism in this chapter, and an advocate of the autonomy conception could consistently deny it. Nevertheless, I think reasons pluralism is plausible. To focus the discussion, I offer an example.⁵

Imagine a group of mountain climbers who, after weeks of struggle, are within a day’s climb of the summit of Everest. They will have to turn back tomorrow. Just after they begin their final push for the summit, they come across a small party of climbers who are huddled together, clearly in terrible need of help. These people, ‘the victims,’ are in their predicament as a result of an extraordinary combination of circumstances, including unusually extreme weather. They need to be helped down to a lower elevation, and their needs are immediate and life-threatening. The first group might realize that, morally, they ought to help. Yet they are

5 The example is based on an event described in Krakauer 1998.

indifferent to moral considerations; they do not care whether they are doing what they morally ought to do. They hurry on to the summit. In doing so, they are being extremely selfish and callous, but, intuitively, this is compatible with their being entirely *rational*.

This example brings out several intuitive points about our ideas of rationality and reasons for action, points that are accommodated by the autonomy conception.

First, the example suggests that a person can be rational in knowingly doing something morally wrong. Kant would disagree, I take it, but I shall not argue against the Kantian view in this chapter.⁶ My focus will be on developing the autonomy conception rather than on arguing against alternatives or exploring the relation between rationality and morality.

One might think that if the climbers believe they ought morally to help the victims, then they have moral values, and thus, on the autonomy conception, they have a self-grounded reason to help. I will argue, however, that an agent's values (in the relevant sense) are not beliefs. Despite their beliefs, the climbers may lack moral values, and thus, on the autonomy conception, they may be rational to ignore any temptation to help the victims. Rationality does not ensure moral virtue. A rational person may have moral values, but she need not.

Second, I think the example supports the intuitive plausibility of reasons pluralism – the view that there are different kinds of reasons, including moral reasons, self-grounded reasons, reasons of etiquette, and so on. And the example suggests, third, that if rationality is a matter of responsiveness to reasons, it is a responsiveness to reasons of a certain kind. Even if there are reasons of etiquette, for example, a rational person might decide not to give them any weight, just as, in the Everest example, the climbers decide to set aside moral considerations. On the autonomy conception, the reasons that a rational agent takes into account in deciding what to do, if she is aware of them, simply in virtue of being rational, are self-grounded reasons – these are (roughly) facts about the impact the agent's alternatives would have on what she values.⁷

Fourth, in the example it seems intuitively plausible both that the climbers rationally ought to continue the climb and that they morally

6 Kant 1785, 50–51 (Ak 447–449).

7 Philippa Foot rejects reasons pluralism in Foot 1978b (161, 168 n. 8). Yet she appears to agree with me that moral considerations are not a source of reasons of a kind that any rational person would take into account in deliberation, if she were aware of them, just in virtue of being rational.

ought to help the victims. Intuitively, just as there are different kinds of reasons, there are different kinds of “ought.” This raises the question of what the climbers ought to do *simpliciter*. I have argued elsewhere that there is not a highest-order normative standard or kind of reason relative to which there is in general something that ought to be done *simpliciter*.⁸ If I am correct, there is no answer to the question of what the climbers ought to do *simpliciter*, but I will not be arguing for this position here. The autonomy conception is compatible with the thesis that the rational action is required *simpliciter* but does not entail it.⁹

One might object that when we are making a decision, we want to know what to do *period*, not merely what to do *rationally*. My account speaks to this concern, but not by assigning a special *metaphysical* status to self-grounded rationality. Rather, as I shall explain, it assigns a *deliberative* priority to self-grounded rationality. Self-grounded reasons play a central role in rational deliberation that is not played by other kinds of reasons, such as moral reasons – unless the agent has the relevant values. One might insist in light of this that, on my account, self-grounded reasons are the only *genuine* or genuinely *normative* reasons. But there are moral considerations that count for and against our decisions, and it would obscure this fact if we denied that moral reasons are ‘genuine’ or ‘normative.’ But this is a side issue. It concerns how best to formulate reasons pluralism, not how best to formulate the autonomy conception.

The autonomy conception has certain theoretical advantages in addition to its intuitive advantages. First, it seeks to *ground* the values standard. Donald Hubin has proposed that rationality consists in conformity to something like the values standard, but he sees no need to ground the standard.¹⁰ Admittedly, it will not be clear why it is an advantage to provide a grounding until I have explained what this involves, but the objective is to explain the normativity of the values standard.¹¹ It is to explain why so-called self-grounded reasons deserve the label “reasons.” I agree, then, with Christine Korsgaard that a principle of practical reason

8 Chapter 9.

9 There is risk of terminological confusion. Some writers think that the English expression “ought” refers, when unqualified, to what I would say “ought rationally” to be done. For example, where Gavin Lawrence speaks of “what the agent ought unsubscripted (that is *qua* rational) to do,” I speak of what an agent “ought rationally” to do. See Lawrence 2002, 120. I think there is no answer to the question of what the climbers ought to do *simpliciter*, but I do think that, in the example as I understand it, they ought *rationally* to continue the climb.

10 See Hubin 2001, 445–468.

11 There are different conceptions of normativity. I discuss some of the varieties in chapter 8.

“needs a normative foundation.”¹² Second, my account offers a kind of realism about rational requirements that is compatible with metaphysical naturalism. On my account, thoughts about rational action and choice are beliefs with naturalistic truth conditions.

In short, I think that the autonomy conception is at the core of an intuitively plausible view about reasons and rationality, which can be integrated into a plausible overall view of deliberation and choice. In the Everest example, if we think the climbers were acting rationally in pressing on to the summit, it is because we understand them as efficiently pursuing their underlying values. Of course, our concern with our *own* rationality is not due, at least not primarily, to a concern to *understand* what we are doing. It is due to a concern, *inter alia*, to do well in governing our lives.¹³ The autonomy conception can explain this. It conceives of rationality as consisting in a kind of success in serving one’s values, and as I conceive of valuing, it is partly constitutive of having values (in the relevant sense) that a person have certain policies for her own behavior. On this picture, as I will explain, rational behavior instantiates a kind of self-government.

3. AUTONOMOUS AGENCY

The literature on autonomous agency is primarily concerned to explain how we can be self-governing despite the variety of causal influences on our actions and despite the possibility of causal determinism. The explanandum in this literature is free intentional action. My project is a different one. I shall set aside the worry about determinism. Moreover, the explanandum in my project is, crudely, intentional action that is *governed* by the agent in a way that she ‘identifies’ with. It is action that is autonomous in a different, ‘thicker’ sense. Cases of *akrasia* or weakness of will count as autonomous action in the thinner sense, if any actions do, but in general they do not count as autonomous in the thicker sense that concerns me, for an *akratic* agent normally does not identify with her action in the relevant way. I shall focus on the thicker notion.

Autonomy is a matter of being ‘self-governing.’ Consider the idea of a self-governing country or state. States are affected in many ways by the

12 Korsgaard 1997, 218, 249.

13 David Velleman proposes that “self-understanding” is the “constitutive aim” of action – in something like the way in which “truth” is the “constitutive aim” of belief (Velleman 2000, 16–24, see 22). My account does not require postulating a “constitutive aim” of action. A thorough discussion of Velleman’s proposal is beyond the scope of this chapter.

actions of other states and by the environment in which they operate, but excessive outside interferences of certain kinds are incompatible with self-government. In a self-governing state, law and public policy originate in decisions of the government, and the government is able to implement its decisions without being subject to certain kinds of interference. By analogy, an autonomous person is someone capable of deciding for herself what to do, and capable of executing these decisions without interference of certain kinds. To be autonomous, one must meet conditions of two kinds. There are 'internal' conditions, including the requirements of being able to make decisions, to form intentions, and to act on one's intentions, and there are 'external' conditions, including the requirement of being free of certain kinds of interference, including coercion and manipulation.

The external requirements of autonomy do not play a role in my account of rationality, for one can deal rationally with external interferences, such as coercion and manipulation. The autonomy conception explains rationality in terms of the internal requirements of self-government; thus, for my purposes here, we can set aside the idea that there are also external requirements.¹⁴

The internal requirements of autonomy are psychological and physical properties and capabilities. First, one must have a 'will.' That is, one must have a structure of beliefs, values, and desires, and the ability to decide how to act and to form intentions to act on the basis of these beliefs, values, and desires. Second, one must have the ability to make one's will effective in leading one to act. That is, one must have the ability to act on one's decisions, to act intentionally. One must have the power to perform bodily movements that, if one is successful, will constitute doing what one has decided to do. The notion of intentional action needs to be explicated, but I believe it is clear enough for my purposes. Autonomous behavior is intentional, in that the agent could in principle give her reason for acting as she does, where her reason is a function of her intention in acting. Of course, intentional action can be irrational, so it is not necessary that the agent's reason for acting be a good reason.

There are important distinctions to be drawn among conative states. Most important for my purposes is the distinction between ordinary desires and intentions. I might desire to eat some ice cream, even though I

14 Our values are shaped by the values of our family and culture, but such influences do not undermine our autonomy. Were it not for such influences, we might not be capable of the kind of planning required for autonomy. See Kymlicka 1995, ch. 5, esp. 82–84. See also Oshana 1998.

intend not to do so, and I might intend to do something, such as to adhere to rigid guidelines about grading, even though, in the ordinary sense of the term, I have no desire to do so. Michael Bratman has proposed that intentions are “planful states” in a way that desires are not.¹⁵ The key point, however, is that there is a distinction.¹⁶

There is also an important distinction between ‘intentions in action’ and ‘prior intentions.’¹⁷ A prior intention can be a specific intention – an intention to do some relatively specific kind of thing at a specific future time. But some intentions govern a kind of action in kinds of situations that I might find myself in on many occasions. An example is my intention to wear a seatbelt when driving. Bratman suggests that we think of general intentions of this kind as ‘policies.’ The idea can be usefully extended to intentions that can be achieved only if a plan is implemented over time, where this involves acting on various more specific intentions that are components of the plan. Policies are intentions that are functionally general in that they guide the formation of various specific intentions over time.¹⁸

To understand the thicker notion of self-governing agency, we need to focus on the idea of *governing* something, understood as a matter of regulating the thing, or exercising systematic control. Autonomous agency is agency that is controlled or regulated by the agent.¹⁹ The ability to have policies or plans is necessary to self-governing agency, so understood; and to the extent that we are self-governing, we shape our lives in accord with our policies or plans. At a minimum, we are not simply driven by our strongest desires, but we select which desires to satisfy and which to treat as ends. There is, for example, the experience, when standing on a tall observation platform, of feeling drawn to jump. To the extent that we are autonomous, we can ignore such desires; we can refrain from treating their objects as goals to achieve. This ability to decide which of our desires to act on is part of what is involved in the ability to plan.²⁰ Beyond this, in planning, we decide among alternative future courses of

15 Bratman 1987.

16 Korsgaard has suggested that “willing an end just is *committing* yourself to realizing the end” (Korsgaard 1997, 245). She understands such commitment normatively, as analogous to “making a promise” (245 n. 60). In my view, to will an end is to form a kind of intention, which is not essentially a normative matter. Korsgaard’s view is disputed in Wallace 2001.

17 Searle 1983, 84–85.

18 Gideon Yaffe pressed me to explain the sense in which policies are general.

19 For a valuable discussion of the agential control of action, see Fischer and Ravizza 1998.

20 Bratman stresses the ability we have to decide whether to take the object of a desire to be an end. See Bratman 1996. Bratman cites Cohon 1993.

behavior, settle on priorities and strategies for achieving our priorities, and of course, if we are self-governing, we act on the basis of such strategies and plans. A person may have various plans or policies, and qualify as acting autonomously in the thin sense, even if she fails in various ways to control or regulate her action systematically on the basis of her plans or policies. She might follow her plans only in a haphazard manner; her decisions might not tend to serve her plans well, or they might not reflect the priorities she has settled on in her planning, and so on. Hence, an agent can be autonomous in the thinner sense without being self-governing in the thicker sense that interests me.

In this thicker sense, I shall argue, a *self-governing* person regulates or controls her actions on the basis of intentions and plans that serve her values, or are at least constrained by her values, such that serving those intentions and plans does not conflict with serving her values. I shall argue that our values are a central subset of our policies, but this is a different point. Here the point is that the values of an autonomous agent constrain the rest of her policies and goals – the ones that do not qualify as values. Hence, the policies and goals that shape her decisions have been shaped by her values. Of course, autonomous agents can act on urges and desires, but they indulge such things within boundaries set by their values. In what follows, when I speak of autonomous agency, I shall intend self-governing agency in this thick sense.

4. MICHAEL BRATMAN ON AUTONOMOUS AGENCY

In introducing the thick notion of autonomous agency, I described it as agency that is controlled and regulated by the agent herself in a way she ‘identifies’ with. I have not yet explained in what sense an autonomous agent identifies with the way she controls actions that are governed by her values. Before proceeding to explain this, it will be useful to consider a different model of self-governing agency, the model that has been proposed by Michael Bratman.²¹ His model is similar to mine, and my proposal builds on his work in the theory of action. The main difference between our views is that Bratman invokes a metaphysical conception of the identity of persons in order to explain the sense in which self-governing behavior issues from the self. I invoke, instead, a nonmetaphysical idea, the idea that an agent can ‘identify’ with a way of controlling action.

21 Bratman presents his account in a series of papers. A useful overview is in Bratman 2004.

Bratman aims to provide a model of the “core elements of autonomy” in at least a significant family of cases of autonomous agency, which he calls cases of “hierarchical self-governance.”²² His model agrees with mine in seeing autonomous action as regulated by a subset of the agent’s policies. But whereas I think the relevant policies are our *values*, in Bratman’s model they are “self-governing policies.”²³ These are second-order policies concerned with the functioning of desires and other conative states in practical reasoning. An example would be my policy of giving weight in deliberation to my desire for safety. Such policies function to guide deliberation. Moreover, because they typically are stable across time, and because guidance by them involves reference to plans and desires the agent has at other times, they play a role in organizing the agent’s life across time. They play this role by means of “continuities and connections” of the kind that, as Bratman reminds us, are central to Lockean accounts of personal identity.²⁴

I need to explain this. At any time that a person is conscious, she is having various experiences, thoughts, feelings, emotions, and so on. As time passes, the person has new experiences, remembers past experiences, anticipates the future, regrets the past, forms plans, adopts goals, learns new things, and so on. These events in the psychological life of a person are related to one another in a variety of ways that go beyond their mere ordering in time. We are not dealing with a kaleidoscopic flux of unrelated events. For instance, there are memories of previous experiences and anticipations of future experiences; there are plans that are formulated at one time and carried out in specific intentions at a later time. A person might have forgotten many childhood experiences by the time she is in middle age, but as a teenager she likely still remembered many of them, and in middle age she might remember much of what occurred to her as a teenager. In this way, events in childhood might be linked to the memories of middle age by an overlapping chain of memories. Given all of this, the Lockean idea is that a person *is* essentially a stream of psychological events and states that is unified by the fact that the events and states in the stream are linked together in a chain by the kinds of psychological continuities and connections to which I have been referring.

22 Ibid., 35–36.

23 Ibid., 43.

24 Ibid., 41–42. Bratman refers to Parfit 1984 (206–208), for a contemporary discussion of a Lockean view of personal identity.

We can now return to Bratman's view. His idea is that self-governing policies function to guide deliberation in a way that involves reference to plans and desires the agent has at other times, including especially future times. These policies therefore organize the agent's life across time by means of continuities and connections of the kind that, in Lockean accounts of personal identity, organize what would otherwise be merely a kaleidoscopic sequence of psychological events into a unified life of a single person. Because of this, Bratman argues, self-governing policies are fitted to constitute "the agent's practical standpoint."²⁵ They are fitted to do this provided the agent is "satisfied" with them. Bratman explains the latter idea negatively: To be satisfied with a self-governing policy P is for P *not* to be in conflict with one's other self-governing policies (and quasi-policies) in a way that tends to undermine the role of P in supporting Lockean continuities and connections.²⁶ Finally, Bratman points out, the policies that figure in his model will be self-referential because they will speak to their own functioning.²⁷ In sum, Bratman holds that, "in a basic case," self-governing agency consists in "the known guidance of practical thought and action by [reflexive] self-governing policies with which the agent is satisfied."²⁸

This is an important and elegant model of self-governing agency, and it deserves a sustained discussion beyond what I can give it here. But I have two worries.

First, I believe it is a mistake to think that autonomous action is typically guided by self-governing policies. I think that what is crucial to self-government is that one's life be governed by one's *values*. In order to make it fully clear why I say this, I need to develop my accounts of values and identity, which I will do in the next two sections of the chapter. But the basic point is simple. I value my safety, for instance, and values of this kind are not self-governing policies because they are 'first-order.' In valuing my safety, I am concerned directly with my safety, not with my conative states or with my practical deliberation. It seems to me, however, that actions that are controlled or governed by first-order values of this kind have at least as good a claim to qualify as autonomous as do actions

25 Bratman 2004, 42.

26 For Bratman, a "quasi-policy" is a "higher-order, policy-like" concern that is not, strictly speaking, a self-governing policy. See Bratman 2000, 57–60 and 49–50; Bratman 2004, 44, 38–39.

27 Bratman 2004, 44.

28 Ibid.

that are controlled or governed by Bratman's self-governing policies. Most of the values of a typical person are not self-governing policies, since they are first-order. It seems to me that self-governance by such values is typical of autonomous agency.

In response, Bratman could point out that seeming first-order values often consist in clusters of policies that include self-governing policies.²⁹ For example, valuing my safety might involve a policy of not permitting my love of adventure to outweigh safety in my deliberation. Still, presumably I would have this policy because I value safety. In general, it seems plausible that self-governing policies are adopted to serve our values. So I see no reason to privilege self-governance by self-governing policies by comparison with first-order policies, such as the simple policy of seeing to my safety. Hence, I think Bratman's model describes a special case of autonomy rather than the typical case.

Second, I think we need a richer account of the agent's endorsement of, or identification with, the elements of her practical standpoint. "Satisfaction" with reflexive self-governing policies is inadequate. To see this, imagine an obsessive person who is obsessed with giving no weight in practical deliberation to her obsessions (including this one). This obsession appears to qualify as a self-referential self-governing policy, and the person may count as "satisfied" with it since its Lockean role may be unimpeded by conflict with any other self-governing policies (or quasi-policies).³⁰ If so, then it may qualify as part of the person's practical standpoint. This seems to be a mistake, because the person neither endorses this obsession nor endorses its having a role in her deliberation. If it has such a role, she might be ashamed that it does. After all, she is obsessed with giving it *no weight*. So I think we should exclude it from her "practical standpoint."

It seems to me, then, that there are two main problems with Bratman's account. In my view, autonomous action is, roughly, action guided by a person's values, where, as I will explain, a person's values figure in her 'identity' in a sense that ensures that she endorses their governing her

29 Bratman urged this point in a helpful correspondence about my criticisms of his view. See also Bratman 2003, 160–163.

30 Her other obsessions may conflict with it, but they typically would not be self-governing policies (or quasi-policies). An obsession with her hair, for example, would be first-order, so it would not be a self-governing policy or quasi-policy, and it might not affect her deliberation, for her obsession not to give weight in deliberation to her obsessions stands in its way. If so, it would not impede the Lockean role of the latter. On Bratman's account, she counts as satisfied with the latter provided that no self-governing policy or quasi-policy undermines its Lockean role. See note 26 above.

action. A person's values can be first-order, so my view avoids the first problem with Bratman's account. Moreover, as I will explain, the fact that a person's values figure in her identity ensures that they are relevantly endorsed, so my view also avoids the second problem.

5. AUTONOMY AND THE IDENTITY OF PERSONS

The life of an autonomous agent is governed in some important sense by the agent *herself*. This may suggest that an adequate account of autonomy needs to invoke a metaphysical conception of what agents *are* – a metaphysical conception of the *identity* of persons. As we saw, Bratman invokes a Lockean conception of the person. I think, however, that an account of autonomy should be neutral among various metaphysical accounts of the person. What we need, to use Bratman's term, is a viable conception of the agent's "practical standpoint," which can be understood, roughly, as the set of psychological states of the person that, when they control her behavior, qualify that behavior as *self-governed action*. The idea of a practical standpoint presupposes that we can identify a person in order to draw a distinction within the person's psychology between different ways that action is controlled, so it assumes that the metaphysics of persons is not at issue. The issue we face is psychological, not metaphysical.

In this section, I shall propose an idea of the psychological identity of persons, and use it to develop a conception of the agent's practical standpoint. The basic idea is that certain of a person's beliefs about her life constitute a whole that we can call her "identity" because of the way that these beliefs ground emotions of esteem, such as pride and shame. Such emotions are related to a person's conception of herself because she 'identifies' in a relevant way with their objects. Philippa Foot has written: "The characteristic object of pride is something seen . . . as in some way a man's own."³¹ I think that a corresponding point could be made about shame and other emotions of esteem. I propose, then, to take a person's identity – her 'self-esteem identity' – to be constituted by (roughly) the set of propositions about her life that she believes and that ground emotions

31 Foot 1978a, 113–114. Foot's dictum would need to be reformulated to deal with cases in which pride takes a propositional object, for a proposition is not viewed as "a man's own." A person may be proud of a painting, or proud *that* she made the painting. Following Foot, we could perhaps say, in cases of the latter kind, that the object of a man's pride is a proposition about an actual or possible state of affairs involving the man – where, she would add, the obtaining of that state of affairs is viewed as some sort of achievement or advantage.

of esteem.³² A person's practical standpoint can be seen as that part of her identity that is concerned, *inter alia*, with plans for her life – plans which are such that her beliefs about her success or failure in accomplishing them ground emotions of esteem.

The use of the term “identity” to express a psychological notion of the sort I have in mind is fairly recent.³³ Korsgaard has proposed that a person's identity is a system of characteristics that the person has and values having.³⁴ But I want to allow for cases in which a person disvalues an aspect of her identity. During apartheid in South Africa, blacks were abused and humiliated on account of being black, and it is likely that many of them disvalued being black. Yet being black is a property that I would want to treat as part of their identity. K. Anthony Appiah has suggested that the identity of a person is a set of “properties important for social life,” which might “matter to their bearers in very different ways.”³⁵ But a person might be tormented by an apparently trivial event that she takes to be quite central to her life even though, and perhaps because, it was not important socially. I would want to treat this as an aspect of her identity.

We should evaluate accounts of identity on the basis of their explanatory usefulness. Different accounts might have different explanatory merits. I initially developed the idea of self-esteem identity in order to cast light on some issues in political philosophy, especially the power and importance of nationalism and patriotism. In this chapter, I will use the idea to illuminate the difference between self-governed action and action that is intentional but that conflicts with our values or is governed only by relatively shallow and perhaps transient goals or desires.

I have suggested that the life of an autonomous agent is governed by her own policies and plans, and, more specifically, by her *values*, which are a subset of her policies and plans. The connection with the idea of identity is, I shall argue, that a person's values qualify as aspects of her identity. This remains to be explained, but if we assume it to be correct, then we can see a connection between governing one's life on the basis of one's values and governing one's life *oneself*. I have introduced the idea that an agent who regulates her life on the basis of her values qualifies as *self-governing*. I

32 I develop this suggestion in detail in Copp 2002.

33 The 1933 edition of *The Oxford English Dictionary* does not give an entry for “identity” with anything like the relevant meaning. In the 1976 supplement to the OED, however, the term “identity crisis” is given, with the first cited usage being 1954. Perhaps, then, we should speak of the “baby boomer conception of identity.”

34 See Korsgaard 1996, 101. Korsgaard speaks of our “practical identity.”

35 Appiah 1994, 150–151.

am now suggesting that she also qualifies as *self*-governing, for her values figure in her identity and constitute her practical standpoint. They ground emotions of esteem that reveal the shape of her self-conception. Moreover, I want to argue, *any* self-governing agent governs her life on the basis of her values. For otherwise, given that her values are aspects of her self-esteem identity (as I will explain), she would be ashamed or disappointed in herself on account of her actions, which would indicate a failure in self-government as she herself sees things. I want to say, then, that an agent is self-governing just in case she governs her life on the basis of her values. Self-governed behavior is not merely intentional and uncoerced. It is regulated by our values, and given that our values are deep psychological features of ourselves that affect our fundamental attitudes toward ourselves, and thereby figure in our identity, self-governed behavior also in this way *expresses* the agent's *identity*. The argument turns on the idea of self-esteem identity, which I shall now proceed to explain.

I need to begin with a brief discussion of the concept of self-esteem.³⁶ Self-esteem is a matter of the degree to which one feels satisfied with oneself on balance, and this is a matter of the degree to which one has a sense of *worth*. A sense of worth should not be identified with a set of beliefs about one's value. A person might believe herself to be valuable, or even to be superior to others, but have low self-esteem, feeling unworthy and insecure. Another person might believe herself to be mediocre, but, despite this, have a solid sense of self-esteem. Self-esteem involves an emotional assessment of oneself, or an emotional stance toward oneself.

A range of emotions is involved. Call them "emotions of esteem."³⁷ On the positive side, a person can feel good about herself, satisfied or comfortable with herself, or have a sense of worth or security or confidence in herself. A person can take pride in various things to which she takes herself to be related in a relevant way. A person can feel enhanced by something. On the negative side, a person might feel worthless or dissatisfied or uncomfortable with herself, or have a sense of insecurity, or lack confidence. She can feel shame, humiliation, or embarrassment. She can feel diminished by something. All of these emotions can bear on a person's self-esteem.

36 In this and the following six paragraphs, I follow the argument in Copp 2002.

37 Gabriele Taylor speaks of "emotions of self-assessment" (1995, 168). Taylor holds that self-esteem is primarily to be understood in terms of pride and humiliation (173). See also Deigh 1995. Deigh relates self-esteem to a range of emotions of the kind I mention (135–139). He links the idea of having disgraced oneself and the idea of shame with the idea of who one is, and he links the ideas of disgrace and shame to self-esteem.

A person's emotions of esteem are grounded in her beliefs, often in a cluster of beliefs, in one of two ways. First, an emotion of esteem might take as its object a proposition that the person believes. For example, a person might be ashamed that she stole a radio. Second, the person might have a belief she would cite, or the propositional object of which she would cite, to explain her feeling. For example, a person might explain being ashamed that she stole the radio by remarking that she knew at the time she was doing something wrong.

Emotions of esteem can be fleeting. It might be a temporary and short-lived fact about me that I feel ashamed of myself for leaving a miserly tip for a waiter. But even if the shame I feel about the tip is short-lived, it might be a relatively enduring fact about me that I feel ashamed when I recall gaffes I have committed, such as leaving a small tip. We are interested in stable and enduring facts of this kind about person's emotions of esteem and their grounds.

Consider, then, the set of propositions about a person that are believed by the person and that ground emotions of esteem in her in one of the two ways given above, in a stable or relatively enduring way. Let us provisionally define a person's self-esteem identity as consisting in this set of propositions. Suppose, for example, that you are proud that you have Greek ancestry, ashamed that you cannot speak Greek, embarrassed that you have a tendency to leave miserly tips, and mortified that you stole a radio. In this case, your belief in certain propositions explains these emotions (or the propositions are the objects of the emotions). These propositions are included in your identity.

We need to amend this account, for there are cases in which people would feel emotions of esteem in various hypothetical circumstances that seem diagnostic of their identity. Imagine a man who is not proud of his Greek ancestry, but who would feel diminished if he somehow came to believe that he did *not* have Greek ancestry. If this is true of him, his having Greek ancestry ought to be counted as part of his identity.³⁸ There are also cases in which a person's emotions of esteem would be affected if, counterfactually, she believed certain things about other people, or about a group or an entity, to which she takes herself to be relevantly related. For example, the man might feel pride if someone else whom he believes also to be Greek won the Nobel Peace Prize. In this case

38 For this purpose, we consider the closest possible world in which the person has the beliefs in question. I am assuming the account of counterfactuals in Lewis 1973.

too, having Greek ancestry ought to be counted as part of the man's identity.³⁹

Here, then, is the proposal.⁴⁰ The self-esteem identity of a person at a particular stage in her life is the set of propositions about herself, each of which she believes, where her belief grounds an emotion of esteem. In some cases, her belief 'actively' grounds such an emotion. In other cases, her belief grounds such an emotion 'potentially' in one of two ways. Either it *would* ground an emotion of esteem if she had certain relevant beliefs about other people or about a group or entity to which she takes herself to be relevantly related, or, if she came to believe its negation, this new belief would ground an emotion of esteem. More formally:

The proposition that she has property F or that she is R-related to an entity E is an element of S's identity during a stage *s* of her life just in case S believes the proposition during *s* and *either*

- (a) this belief grounds an emotion of esteem in a stable or relatively enduring way during *s*, or
- (b) it would do so, if S had certain beliefs about E or about other people whom she believes to be F, or to be R-related to E, or
- (c) if S were to come to believe during *s* that she is not F, or that she is not R-related to E, then, other things being equal, this belief would ground an emotion of esteem in a stable or relatively enduring way during *s*.

For present purposes, the important point is the connection between self-esteem identity and self-government. Suppose that you have planned your life around various projects, such as raising a family, excelling in your career, being fair in your dealings with others, and so on. The degree to which you are content with yourself will be grounded, among other things, in the degree to which you believe you are finding success in these projects. Since self-contentment is an emotion of esteem, this means that the fact that these are your projects and that you are succeeding in them is an aspect of your identity. Moral commitment and moral character are also entangled with your identity. If a person subscribes to a moral principle, attitudes of shame, guilt, or contentment would be grounded in her record of compliance with the principle. Indeed, the fact that she subscribes to the principle likely would be an aspect of her identity, for she likely would feel ashamed if she came to believe that she does

39 Nomy Arpaly helped me with this kind of example.

40 This paragraph is taken with some modifications from Copp 2002.

not actually subscribe to the principle. And the things a virtuous person believes about her moral character also normally ground emotions of esteem such as pride or shame, either actually or potentially. A virtuous person would see herself as honest, as not manipulative, and so on, and she would feel ashamed or diminished if she came to see herself as dishonest or as manipulative. Hence, her identity would normally include propositions such as that she is honest and nonmanipulative.

Similar considerations show that a person's *values* are aspects of her self-esteem identity. If a person values honesty, she would tend to feel ashamed of herself if she were to realize she had acted dishonestly, for instance, and she would also feel ashamed of herself if she came to believe that she is not an honest person. Hence, the fact that she values honesty will be an aspect of her self-esteem identity. Nor is this merely likely to be the case. It seems to me that a person who values honesty must have a tendency to feel ashamed if she thinks she has acted dishonestly or if she comes to believe she actually is not honest. This is why the presence or absence of such emotions is evidence of a person's values.⁴¹ In general, then, on my account of valuing, and on my account of identity, a person's values are aspects of her identity.

There is a technical difficulty that needs to be addressed. A person's identity is a set of propositions she believes. The difficulty is that, on my view, as I will explain more fully in the [next section](#), a person's values are policies or general intentions rather than beliefs. Moreover, such a policy can be merely implicit in the sense that one might not be able to formulate it and one might not follow it self-consciously.⁴² Hence, if I want to maintain that a person's values are 'aspects' of her identity, it appears I need to say what proposition a person believes when she values *V* such that (a) it qualifies as an element of her identity on my account of self-esteem identity, and (b) it concerns *V* in a relevant way or is appropriately related to *V*. Suppose, then, that you value honesty. In this case, I think you would believe that you have a policy of being honest. This belief might be merely implicit, in that you might not have it consciously in mind, and it might be rather inarticulate, in that, for example, you might not think of yourself as having a 'policy' rather than simply a desire to

41 Aristotle suggests that our character is revealed in our feelings. See Aristotle 1998, book 2, ch. 3, 1104b3ff. I owe this reference to Elaine Sternberg.

42 We presumably have a policy of speaking grammatically, but we might not be able to formulate the rules of grammar and might not follow the policy self-consciously. The policy is 'implicit' in this sense. Frank Jackson describes grammar as an "implicitly understood theory" (2003, 569).

be honest.⁴³ Nevertheless I think some such belief would be present. Jay Wallace has argued that our intentions are ordinarily “fairly accessible to consciousness,” given the role they play in shaping our deliberation. He argues, in fact, that someone who does not believe that she intends to do x “cannot really be described as having the intention to do x.”⁴⁴ The point, of course, is that policies are general intentions. My proposal, then, is that if a person values V, the proposition that she has a policy of pursuing V is an element of her identity. Indeed, a cluster of beliefs regarding V will normally be included in her identity, including especially beliefs about her success or failure in serving V. For example, if she believes she missed an opportunity to serve V, she will tend to be disappointed in herself. A person’s values are revealed by the emotions of esteem that she feels or tends to feel, and her values are embedded in her psychology in virtue of their connection with such emotions. Given all of this, I think it is appropriate to speak of a person’s values as being ‘aspects’ of her identity.

A person *endorses* the role of her values in governing her actions in that she is content with herself when she deliberates and acts in accord with them, ashamed to fail to do so, and so on. To see the importance of this, notice that a person with an obsession might be ashamed of her obsession and ashamed to indulge it. In such a case, the fact that she has the obsession would be an aspect of her identity, but we would not want to count indulging it as an instance of self-government. This is for two reasons. First, the obsession is not an aspect of the person’s identity in the right way. Control of one’s deliberation and actions by one’s values is control by an aspect of one’s identity that is endorsed in one’s identity, where such control is also endorsed in one’s identity. Obsessions are not normally endorsed in this way, nor is their role in one’s life. Second, an obsession is not normally a policy, and the compulsive nature of an obsession subverts the control of one’s actions by policies. In discussing Bratman’s view (in section 4), I gave an example of an obsessive policy, but the policy in that example was in part a policy of giving itself no weight in deliberation, so the agent did not endorse its playing a role in her deliberation.

Let me then return to the argument that began this section. If an agent regulates her life on the basis of her values, she qualifies as

43 Jay Wallace speaks of beliefs that are not “explicitly and articulately present to the consciousness of the agent” but that “are implicit in the agent’s understanding of their situation” (2001, 21 n. 52).

44 Ibid., 22.

self-governing, for her values figure in her identity. Her actions express her self-conception. And she qualifies as *self-governing*, for she endorses control of her actions by her values and endorses her values. Moreover, if an agent is *self-governing*, she governs her life on the basis of her values, for otherwise she would be ashamed or disappointed in herself on account of her actions, which would indicate a failure in self-government as she herself sees things. Hence, as I claimed, an agent is (thickly) *self-governing* just in case she governs her life on the basis of her values. Her values constitute her practical standpoint.

6. VALUES AS POLICIES FOR ACTION

I have argued that (thickly) autonomous agency is agency that is regulated by the agent's values. And I have suggested that our values are a kind of policy, where a policy is a kind of general intention. I now need to defend this suggestion. I suggested that our values figure in our *identity* in a way that ordinary policies and plans need not. I want to argue that, because of this, and because we endorse control of our actions by our values, to the extent that we are *self-governed*, our actions are shaped and constrained by our values. To make the argument go through, however, I need to show that values can be understood as things that *can* regulate behavior, and to achieve this, I want to argue that values are a kind of policy for action. I also need to argue that our values are basic or fundamental policies that, to the extent we are *self-governed*, constrain the more occasional and instrumental plans and ends that shape action. The fact that our values are *intrinsic* policies helps to support this view, as I shall explain, as does the fact that our values figure in our *identity* in a way that ordinary policies and plans do not. The latter fact means that there are emotions of esteem to back up the controlling and regulating function of our values. It means that the engine of self-esteem helps to regulate the actions of a *self-governing* agent.

Let me begin to explain why I hold that our values – or our values ‘for action’ – are a kind of policy. We need to distinguish, of course, between the things that a person values and her state of valuing those things. Consider, for example, what would be involved in valuing honesty, friendship, and so on. I think it is plausible that a person who values honesty must have a policy of being honest. She might sometimes fail to be honest, of course, but honesty must be her policy. It is not enough, for example, that she be honest merely as a result of finding honesty in her best interest from time to time. T. M. Scanlon suggests that to value friendship

is among other things to have a policy of being a ‘good friend,’ of being loyal to friends, of being concerned with their interests, of spending time with them, and so on. It is also to have a policy of seeking to have friends, to keep the friends one has, and to want those we care about to have friends of their own.⁴⁵ Thus, I think that to value honesty and friendship would be, *inter alia*, to have a set of rough-and-ready policies regarding truth-telling, spending time with friends, and so on. In general, I think, to value something ‘for action’ is, *inter alia*, to have a kind of policy with respect to it.

More is involved in valuing than having policies. The policies in question must be relatively stable. Our values do not switch on and off like lightbulbs. They are policies of and for temporally extended beings. Normally, moreover, a person who values something is satisfied with the relevant policy and does not regret it or desire to lose it.⁴⁶ Of course, we can imagine circumstances in which a person who values honesty would regret her honesty. Perhaps honesty has cost her a friend, for example. A person can have a kind of alienation from her values, perhaps viewing herself as naive for valuing honesty the way she does.⁴⁷ Nevertheless, such regret or alienation cannot go too far. For if a person values honesty, she would tend to feel shame or guilt or disappointment in herself, or regret, if she failed to be honest on a given occasion. Some of our policies are not like this. For example, even if I have a policy of reading the newspaper every morning, I would not feel disappointed in myself if, on some occasion, I listened to the news on the radio instead. This shows, I think, that I do not *value* reading the newspaper as such. Perhaps I value getting the news, but if so, then I would feel disappointed in myself if I failed to do anything at all to get the news. Our values for action, therefore, are part of our self-esteem identity, as I have already explained.

I now want to add that I have in mind things that we value *intrinsically*. Our values are ‘intrinsic policies.’ This idea is easiest to characterize negatively, as follows. If a policy is intrinsic, we do not have it merely because we think that carrying it out will be or may be instrumental to carrying out other policies that we have or achieving other things that we want.⁴⁸ Instrumental plans and ends rest on other ends, but intrinsic policies do

45 Scanlon 1998, 88.

46 Here I am influenced by Bratman’s views about self-governing policies (Bratman 2004, 41–45).

47 Kadri Vihvelin suggested that a person can be alienated from her own values.

48 Cf. Scanlon 1998, 79.

not. For example, my policy of wearing a seatbelt is not intrinsic, since I have that policy only because complying with it contributes to my safety. But my policy of seeing to my safety is intrinsic. I do not have it merely for instrumental reasons. Thus, given that our values are intrinsic policies, they are more basic in the government of our lives than our more occasional and instrumental plans and ends.

I propose, then, that for a person to value something ‘for action’ is at least in part for her to have an intrinsic policy of choosing or acting in relevant ways, a policy that is relatively stable, that on the whole she is content to have, a policy compliance with which affects her emotions of self-esteem and which is therefore partly constitutive of her identity, and a policy whose role in governing her actions she endorses.⁴⁹ This account could be taken as a stipulative, but I think it meshes with an ordinary understanding of what is involved in valuing.

One might object that there are cases in which valuing does *not* seem to involve having policies for action, at least not centrally. These are cases in which we value things we believe we cannot affect in any significant way. For example, I might value the accomplishments of Aristotle.⁵⁰ I realize I can do nothing to affect the accomplishments of Aristotle, so my valuing them would seem not to involve my having any policies.

In response, I propose that, even in cases of this kind, a person must have relevant *dispositions*. Suppose, for example, that a person values the pristine environment of Ellesmere Island. If she came to know that the island is threatened by pollution, she surely would have a tendency to support calls for conservation and the like. Having such dispositions is the kind of thing involved, *inter alia*, in having a policy of the relevant kind. Moreover, the problem cases seem to be ones in which a person values an instance of a kind that she values, where valuing the *kind* involves having a relevant policy. For example, a person who values Aristotle’s accomplishments may value them as an example of philosophical accomplishment, and if so, she would have corresponding policies, such as a policy of encouraging, supporting, and applauding such accomplishments. But the important point is that there is such a thing as having a policy, and valuing something ‘for action’ does involve having relevant policies.

49 In Copp 1995, I explained valuing in a slightly different way, in terms of subscribing to a standard, but there is a close connection between subscribing to a standard and having a policy (177–178, 84, 87–88). I proposed a different kind of account in Copp 1993; it was criticized in Gilbert Harman (1993).

50 Various people pressed me to discuss examples of this kind.

My proposal, after all, is to explain rationality and self-grounded reasons in terms of the values standard, and the values standard is to be understood as concerned with our values for action, values the having of which involves, *inter alia*, having relevant policies. Our 'values for action' are a subclass of our values. (If this is doubted, as I said, we could take my usage to be stipulative. The values standard would then be concerned with policies of the relevant kind.)

Various other proposals have been made about what is involved in valuing something. Scanlon holds that to value something is to take oneself to have *reasons* for holding positive attitudes toward it and for acting in certain ways in regard to it.⁵¹ There are several problems with this proposal. First, it makes valuing something more of an intellectual matter than seems plausible, since it seems to imply that one must have the concept of a reason in order to value anything. Second, it seems that a person's values might run contrary to her beliefs about reasons. A person might take herself to have (aesthetic) reasons to hold positive attitudes toward listening to classical music, but she might not value listening to it. She might not enjoy it, and she might have a policy of avoiding it.⁵² If she enjoyed jazz and listened to jazz on a regular basis, then it would be more revealing, I think, to describe her as valuing jazz rather than classical music, regardless of whether she would agree that she has (aesthetic) reasons to listen to jazz. Finally, and most important, Scanlon's proposal is not open to me. He aims to explain valuing in terms of taking oneself to have a reason, but one thing I am trying to explain is what a person would be thinking, in thinking she had a reason. I need an account of valuing that does not use the notion of a reason because my goal is to explain self-grounded reasons in terms of the attitude of valuing.

One might suggest that to value something is to *believe* it is valuable. It seems to me, however, that this proposal also makes valuing more of an intellectual matter than is plausible. A person might have no *beliefs* about the value of listening to music yet still *value* listening to jazz. The proposal also runs into difficulty with pluralism about the bases of value. A person might believe that listening to classical music is aesthetically valuable yet think that listening to jazz is valuable as a source of pleasure. Her values might track the latter beliefs rather than the former. Finally, a person's values might run contrary to her beliefs about what is valuable. She might believe that listening to classical music is valuable, yet she might strongly

51 Scanlon 1998, 95, also 87–100.

52 This example was suggested to me by Richard Schubert.

dislike it and have a policy of seeking out jazz. We *could* say that she “values” listening to classical music, simply to report her belief, but in most contexts, I think it would be misleading to say this.

It is plausible, nevertheless, that valuing something (in my sense) typically goes hand in hand with believing the thing is valuable or good, in at least some respect. Such belief is not necessary, however. In many cases, valuing something (in my sense) *precedes* believing that the thing is valuable. For example, in most cases, I believe, people come to value having a family without first having beliefs about the value of a family. There is perhaps a biological basis to our valuing having a family, which would help to explain this. For my purposes the important point is that, as the example suggests, a *belief* in the value of something is not invariably prior in the order of explanation to *valuing* the thing (in my sense). Moreover, the example shows that it need not be *irrational* for a person to value something in a case in which the fact that she values it (in my sense) is not explained by a belief that the thing is valuable. And it shows as well that it need not be irrational to value something (in my sense) *without* believing it is valuable. We are not irrational to value having a family in the years before maturity brings an understanding of why having a family is valuable.

One might object, however, that it *would* be irrational to value something (in my sense) while believing that the thing has *no* value *at all* or that it is not good in *any* respect. It might seem that there would be a kind of incoherence in this combination of valuing and belief. A person is content with herself when she acts in accord with her values, but it might seem that it would be incoherent to be content with oneself for acting in accord with a policy if one believes that it, or its object, has no value at all. In response, I concede that emotions of esteem depend on one’s beliefs in subtle ways.⁵³ Yet, despite this, a person can value something that she believes not to be good in any respect, and can do so, I think, without irrationality. Someone might value listening to jazz even if she believes there is nothing good at all about listening to jazz. When asked why she spends so much time listening to jazz, she might reply that she loves it even though she can’t explain why. I am not convinced that there would be an incoherence in her state of mind or an irrationality in her behavior.⁵⁴

53 This idea has been addressed by many writers on the emotions. See Foot’s brief groundbreaking discussion in Foot 1978a, 113–114.

54 This point needs more discussion. Steven Wall helped me to see the need to mention it.

My proposal, then, is that to value something is, *inter alia*, to have a policy. Bratman has suggested that to have a policy is to have a general intention governing kinds of action in kinds of situations.⁵⁵ If this is correct, it is no surprise that deliberation that begins with our values can lead to action. If I have a policy of seeing to my safety, and if this involves having a general intention, then we can see why I form an intention to wear a seatbelt on getting into a car since I believe that wearing a seatbelt serves my safety. If to value something is to have a general intention, then in reasoning about how to achieve what we value, we will form more specific intentions and, when appropriate, form the intention to act, other things being equal. In this way, the account of valuing that I recommend makes transparent the relation among valuing, deliberation, and action. This is a theoretical advantage of the account.

In this section, I have been defending my view that our values are central to the autonomous regulation of behavior. First, I explained that our values are a kind of policy. Second, I argued that because they are intrinsic policies, to the extent that we are self-governed they govern our formation of specific instrumental plans that then constrain and guide our actions. Moreover, given this, and because of the way our values are components of our identity, cases in which we govern our actions by our values qualify as cases of self-government.

7. RATIONALITY AND VALUES

According to the autonomy conception, the concept of rationality is closely related to the concept of autonomy. Rationality is a matter of doing well at self-government (understood in the thick sense) and at securing the requirements of self-government. Controversy would center on what governing oneself well consists in. I have been arguing that governing oneself (understood in the thick sense) is basically a matter of living in accord with one's values. If so, governing oneself well must consist basically in doing well at living in accord with one's values (and securing the requirements of doing so). This then is what I propose rationality to consist

55 Having such a policy could be understood as a matter, *inter alia*, of intending to comply with certain relevant standards or norms. This suggestion brings my discussion of valuing into contact with Allan Gibbard's moral psychology, which postulates the state of norm acceptance, and the psychology I have proposed, which postulates the similar state of subscription to a standard. See Gibbard 1990 and Copp 1995.

in. Ignoring certain qualifications, I propose the following standard, which I call the “values standard”:

A person is to serve her values as well as she can, overall and in aggregate, and in situations where more than one alternative would maximally contribute to serving her values, she is to serve her (other) intrinsic goals as well as she can, overall and in aggregate.⁵⁶

To a first approximation, I hold that this is the standard that agents comply with insofar as they are rational – assuming that they have approximately accurate beliefs about what they value and about how to achieve what they value. It suggests the following principle:

Agent S is rationally required to do A in circumstances C just in case either (a) doing A in C is the action that would best serve S’s values, overall and in aggregate, or (b) doing A is one of a group G of actions open to S in C, each of which would serve her values, overall and in aggregate, better than anything else she could do, and, of all the actions in G, doing A in C would contribute most to serving her (other) intrinsic goals, overall and in aggregate.

One might have various objections to this proposal, including the objection that a person’s values can be pernicious. I will not be able to address every objection, but I will discuss those that seem most important. Before doing so, however, I need to mention four caveats.

First, there are constraints on the overall structure of a person’s values, goals, and beliefs. To the extent that a person is rational, her values and beliefs are coherent with one another, and her (other) intrinsic goals are coherent with her values and constrained by them and by her beliefs, as well as coherent among themselves. Whether an agent is rationally required to attempt to achieve coherence in a given case depends on the costs and benefits of doing so, given the time and effort that would be required and given the agent’s values and situation.⁵⁷ Second, a rational person serves her values as well as she can in light of the information

56 A person’s goal is intrinsic if the person does not have it merely because she thinks that achieving it will be or may be instrumental to achieving other things that she wants. The formulation in the text ignores some of the caveats I mention in what follows. I need to leave the notion of ‘serving’ one’s values unexplicated. But see note 4 above. Important complexities lurk under the surface here. Philip Pettit has proposed that “serving” something one values might consist in “promoting” it or in “honoring” or “expressing” it. In some cases, it may seem more appropriate to “honor” something we value than to “promote” it. See Pettit 1991.

57 There are complexities here that I must pass over. Cf. Rawls 1971, 143. See Scanlon 1998, 121–122.

available to her. She can be led by misinformation to believe that some action will serve her values when it will not, or to think that something will not serve her values when it will. Yet when her beliefs are reasonable, she may be acting rationally in acting on those beliefs.⁵⁸ To be sure, a rational person seeks information that is relevant to her capacity to serve her values and goals. She seeks such information to the extent that she is reasonable to believe that she should, given her values, her epistemic standards, and her epistemic situation.⁵⁹ And she assesses what to believe in a given epistemic situation in light of her epistemic standards or values.⁶⁰ Third, a rational person assesses her values and goals in light of new information when she believes she has reason to reconsider them in light of the new information. Fourth, where a person's serving her values (or goals) would put at risk her ability to sustain herself as an autonomous agent, rationality permits her to choose to sustain herself as an autonomous agent, by seeing to it that she is able to meet her basic needs.⁶¹

58 Hence, the autonomy conception makes room for both kinds of error that Gavin Lawrence thinks any "proper theory" would accommodate (2002, 121).

59 A person's "epistemic standards" are norms she accepts that pertain to the rational formation of belief. An example is the rule to believe something only if the evidence or supporting reasons on balance make its truth significantly more likely than its falsity. By a person's "epistemic situation" – with respect to a given proposition – I mean her situation as it bears on the reasonableness of believing the proposition, given her epistemic standards.

60 I pass over the problem of grounding epistemic standards. It is plausible that epistemic standards are appropriately grounded when, roughly, conformity with them would serve us well, in the typical circumstances in which we human beings actually find ourselves, in attaining true beliefs and avoiding false beliefs. On the account of practical rationality I am proposing, it is possible, even if unlikely, that a person who qualifies as practically rational might have epistemic standards that are not grounded in this way. For her epistemic standards may be such that conformity with them serves *her* well in pursuing what she values even though it is not the case that conformity with them would serve human beings well, in typical circumstances, in attaining truth and avoiding falsehood. See Kitcher 1992, 53–114, e.g., 63. Cf. Rawls 1971, 397.

61 Meeting one's needs is a precondition for an autonomous life, so this qualification fits within the autonomy conception of rationality. I am not certain how the qualification should be worded. It should perhaps be restricted to emergency situations. The idea would be that rationality permits securing one's needs in emergency situations, even if doing so is contrary to serving one's values. For example, even if a rescue worker values his work, he may be rational to balk at going into a burning building when the risk to his life is excessive. By an emergency situation, I mean a situation with two key characteristics. First, it is reasonable for the agent to believe that serving her values would undermine or put seriously at risk her ability to sustain her status as an autonomous agent, not in a merely temporary or minor way, such that she could later compensate for neglecting her needs, but in a permanent and decisive way. And second, it is reasonable for her to believe that sustaining her status as autonomous would permit her to carry on in the future to serve at least some of her values. A version of the values standard that is qualified in the relevant way could be called the "needs and values standard."

Even if we set aside these caveats, however, there is more to rational agency than merely acting *in accord with* the values standard, since this could happen by chance. Rationality is a matter of *governing* oneself well, and how well a person is doing at governing herself depends on the way in which she makes her decisions and not merely on what she decides to do. Setting aside various complexities, the basic idea is that rational behavior is *guided* by one's *values*. Such guidance seems to require a person to have at least approximately accurate beliefs about the content of her values, and to act in the belief that her action 'makes at least as much sense,' given her values, as would anything else.⁶² In some cases, an agent is guided by her values in an explicit way. She deliberates about what to do, beginning with her values, considering how best to effect them, and finally reaching a decision about what to do. In many cases, however, an agent acts without consciously going through any reasoning. If she is guided by her values, there is something she values such that she does what she does *because* she thinks that her action *makes sense* in light of that thing. But she normally would not think that the key property of the thing she values is just that she values it.⁶³ She need not be guided by the values standard as such.⁶⁴ Thus, ignoring certain complexities:

In a circumstance where agent S is rationally required to do A, and where S has approximately accurate beliefs about what she values, and about how to serve or achieve what she values, S is fully rational in acting only if (1) S does A intentionally, and (2), where V is something S values, S does A for the reason that, among other things, S judges that doing A will best serve V, and (3), where $V, V_1, \dots, V_n$, is the set of things S values, S judges that doing A makes the most sense given $V, V_1, \dots, V_n$.⁶⁵

62 The locution "makes sense" is used in Gibbard 1990.

63 For a related point, see Pettit and Smith 1990.

64 That is, a rational agent need not use the values standard as a 'decision procedure.'

65 I ignore the possibility of situations in which the agent judges that several actions would serve her values equally well. Where agent S is rationally required to do A because doing so will best serve her values, our assessment of the rationality of S's performance should be sensitive to, among other things, (1) whether S does A intentionally, (2) whether S does A for the reason that it will best serve her values, and (3) whether S believes that doing A makes the most sense given her values. In ordinary *akrasia*, the agent satisfies (3) but fails to do A. She may succeed, however, in serving something else that she values or takes as an end. Some cases of this kind will qualify as cases of "cleverness," as Wallace uses the term (2001, 1). In inverse *akrasia*, the agent succeeds on count (1), and perhaps also on count (2), but believes that something other than A makes the most sense. In another kind of case, the agent succeeds on counts (1) and (3) but acts for the wrong reason. See Arpaly 2000.

The reasons *for which* she acts should correspond to the self-grounded reasons that there are.

This suggests the following picture: A rational person – a person who complies with the values standard – governs herself on the basis of her values, making decisions that serve her values well given the priority she assigns to different values, and given the information she has. In situations in which all of her options would do equally well at serving her values, her values give her discretion, and she can do what will best satisfy her other intrinsic goals. She seeks information she needs in order to govern herself well, and she assesses the information and decides what to believe in light of her epistemic standards. She also seeks to sustain her ability to govern herself in this manner, at least insofar as doing so is compatible with her values, given her information. Rationality permits her to reconsider what to value, even if it calls for doing this in light of other things that she values. If a person values a life spent in the great outdoors, say, then, other things being equal, she lives this way despite the temptations of the city. But if she has values that would better be served by an urban life than by life in the woods, she might face difficult decisions. She might decide that her other values argue against the outdoors life, and she might give it up. Eventually she might cease to value it. This kind of change of values, in the interest of furthering one's ability to serve one's values, can be entirely rational.⁶⁶

8. THE OBJECTION FROM UNGROUNDED ENDS

One might object to the values standard on the ground that a requirement to serve one's values can be plausible only for values that are themselves supported by reasons. People can have irrational and immoral values. Stephen Darwall writes: "If one had no reason to adopt A (or worse, reason not to do so), then maybe [instead of *serving* A], one should give up A." He writes: "From the facts that one has adopted A as an end and that B is a necessary means to A, it does not follow that one ought or has reason to B." It follows at most that one is required *either* to take the necessary means to A *or* to give up A.⁶⁷ Call this the "objection from

66 A similar point is made by Hubin 2001. David Schmidtz points out that having certain "final ends" can be a means to satisfying certain ends regarding our ends (1995, 61, 77).

67 Darwall 2002, 7. See also Broome 1999; Greenspan 1975; and Korsgaard 1997, 252. Korsgaard writes: "Hypothetical imperatives cannot exist without categorical ones, or anyway without principles which direct us to the pursuit of certain ends, or anyway without *something* which gives normative status to our ends" (250).

ungrounded ends.” In effect, the objection is that because people can have irrational and immoral values, compliance with the values standard cannot be a requirement of rationality.

I agree, of course, that the values standard needs to be grounded, and I shall propose a strategy for grounding it. This proposal will be the heart of my reply to the objection. There are, however, some preliminary responses to consider, the most important of which turns on a distinction between rationality and what I will call “wisdom.”

It will be useful to begin by considering a related objection, the objection that the values standard needs to be amended to take account of *irrational* values. I agree that a *set* of values can be irrational – the values of a rational person must be coherent with one another. But I deny that a value can be irrational, just as such, merely in light of its content.⁶⁸ One might object that since I hold that the values standard is grounded in autonomy (in a way I shall explain), I ought to admit that counter-autonomous values – values the serving of which is not compatible with one’s autonomy – are irrational. Perhaps it is irrational to have a policy of always deferring to one’s parents.⁶⁹

I do not want to accept this proposal. A person who has a policy of always deferring to her parents may be doing well at governing herself in light of this policy. I do not want to say that such a person is automatically irrational. For I want to distinguish objections to her values from criticisms of her success in governing her life in accord with her values. I say that objections to a person’s values do not speak against her rationality. I do not want to view people from cultures that do not value autonomy as irrational if they merely have counter-autonomous values. To do so would blur the important distinction between rationality, understood as a matter of how one governs oneself, and ‘wisdom,’ understood as a matter of having values that are morally and otherwise acceptable. The objection from ungrounded ends also blurs this distinction, since it denies the rationality of serving values that are not based in reasons.

The intuitive basis of the distinction between rationality and wisdom is suggested by the Everest example. The climbers were not *wise* to give such importance to climbing Everest, but, despite this, we see them as having been rational in their pursuit of their goal. The values of a rational person may be subject to a variety of criticisms; they may be immoral or impolitic

68 I ignore values that are logically impossible to satisfy. For useful discussion, see Hubin 1991, esp. 24–26.

69 I owe this objection to Steven Davis and Melinda Ammann.

or self-aggrandizing or foolish. A good navigator may not be navigating toward a good destination. Of course, nothing turns on my choice of words to mark the distinction. It would not be a misuse of English to call the climbers irrational for giving such importance to climbing Everest, or to say it would be irrational to have a policy of deferring to one's parents. The important point is that there is a distinction between evaluating how well a person does at governing herself, given her values, and evaluating a person's values. I am concerned with evaluations of the former kind.⁷⁰

Let me return, then, to the objection from ungrounded ends. I agree with Darwall that the fact that one has adopted an end does not entail that one has a reason to serve it, and I agree with Wallace that there is "no genuine requirement to take the means that are necessary for realizing ends that one merely happens to desire."⁷¹ First, the values principle is concerned with serving values, not 'ends.' A value is not an end that one merely happens to desire. A person might be rationally required *not* to pursue an end, if doing so would conflict with serving her values. Thus, the fact that I have a given *end* does not mean that I have any reason to serve it. Second, I can agree that it does not *follow* from the fact that I value something that I have reason to serve it. The values principle is a substantive principle that needs to be grounded.

To be on target, then, the objection needs to be reformulated. The objection is that I have not explained why or how it can be that "ungrounded" *values* are a source of reasons. Michael Bratman has pointed out that this objection may be especially difficult for me, given my view that values are general intentions or policies.⁷² *Specific* intentions do not seem to be a source of *basic* or *undervived* reasons.⁷³ If I form the intention

70 Gavin Lawrence discusses a view he calls the "end-relative account," which is similar to the conception of self-grounded reason that I am defending. He says that on this account "there is simply no question of whether an end is good or worth pursuing" (2002, 115, 128). I am claiming that it *is* possible to evaluate a person's values but that this is a different matter from evaluating the rationality of her actions. Korsgaard says that "the instrumental principle," which is similar to the values standard, cannot stand without "*something* which gives normative status to our ends." Yet she concedes that this status may be very thin. She allows that a "heroic existentialist" might endorse his ends for no "further reason," without thinking his ends are good. See Korsgaard 1997, 251 n. 74 and 252. I have said that an agent 'endorses' her values and the role of her values in governing her actions. This is not to say, however, that an evaluation of the rationality of an agent's actions in terms of the values standard requires or involves an evaluation of the agent's values.

71 Wallace 2001, 1.

72 Bratman pressed this objection in personal correspondence.

73 Scanlon 1998, 70. Bratman stresses that once one has formed an intention to A, one is under certain rational constraints with respect to this intention, such as the constraint not

to go to the gym, this gives me reason to take my gym clothes with me, but it does not give me a new reason to go to the gym. General intentions may seem to be similar: We adopt general intentions on the basis of reasons, but perhaps they are not sources of new reasons. The objection is that if values are just a kind of general intention, I need to explain why or how they can be a source of basic reasons when other intentions are not.

A fundamental response to these objections will have to wait until I discuss the grounding of the values standard. But I want here to bring out the plausibility of the idea that values can be a source of basic reasons – even if values are not themselves grounded in reasons, even if they are simply a kind of intention, and even if specific intentions are not a source of basic reasons.

Nevertheless, I think that a specific intention *can* be a source of a basic reason. If, on a whim, I form the intention to smell a nearby rose, this ordinarily would give me a reason to smell it, assuming that doing so would not conflict with my values. This reason would be basic, in that it would not be derived from any other reason.⁷⁴ But suppose I form an intention to act in a way that, I realize, conflicts with doing what would best serve my values. Suppose, for example, that I decide on a whim not to wear my seatbelt. Since I value my safety, I have reason *to* wear the seatbelt, and – in the given example – *no* reason *not* to wear it. Intuitively, even if I have *no* reason to value my safety, my value gives me a reason to wear the seatbelt, but my whim gives me *no* reason *not* to wear the seatbelt. A decision or intention motivated by a whim does not give a person a reason to act contrary to her values. Two factors seem crucial to explaining this.

First, values are intrinsic. Ordinary policies and intentions typically are formed because we see acting in accord with them as means to serving our values. This is why action in accord with our values is fundamental to governing ourselves in a way that action in accord with ordinary policies (and whims) is not. Second, and more important, our values are aspects

to change the intention without reason. But this is not to say that one acquires a new reason to A merely in virtue of forming the intention. See Bratman 1987.

74 But would the reason be derived from a reason given to me by the whim or urge to smell the rose? There is a temptation to say so. However, an urge or whim can instead be viewed as an occurrence that is not in itself reason-giving. A person can decide whether to indulge it or not, and either way, arguably, she makes no mistake, other things being equal. If she forms the intention to indulge it, then, arguably, she thereby acquires a reason to do so. On this view, a person who stops to smell a rose on a whim may be no more responsive to reasons than would have been the case had she not stopped. Yet, of course, in the latter case, we might doubt her wisdom in eschewing an innocent opportunity for enjoyment.

of our identities in a way that our ordinary intentions, policies, ends, and whims are not. Failing to live in accord with our values negatively affects our self-esteem. In this sense, a person's values are linked to her sense of self. The self-esteem of a person who sees herself as having failed to comply with her own values is shaken to some degree by a sense of failure and disappointment in herself or a sense of shame or regret. A knowing failure to serve one's values is a kind of self-betrayal.⁷⁵

Let me return, then, to the original objection, reformulated as the objection from ungrounded values. The objection is that a person who values A without having any reason to value A is rationally required, at most, *either* to serve A *or* to give it up; she is not rationally required to serve A. But consider the alternatives. A person does not have the *option* of giving up a value in the way that she has the option of giving up an ordinary goal, such as to smell a rose. We cannot change our identities at will, and our values are embedded in our identities. Of course, our values can *change* – even as a result of decisions we make. A person might rationally try to change her values in light of other things that she values. But we cannot *decide* to give up a value, or even an ordinary end, in the way we can decide what to do. At the point of action, our values are set. Hence, if a person with a value is rationally required either to give up the value or to serve it, then since she lacks the option of giving it up, it is plausible that she is required to serve it. It is plausible that to the extent that she is rational, she will govern her action in accord with it.

The point here is not simply that we cannot change our values at will. The reason we cannot change our values at will is, in part, that they are embedded in our identities; they are grounds of our self-esteem and aspects of our self-conception. Otherwise they would not function as they do, as compasses in our lives.⁷⁶

9. GROUNDING A CONCEPTION OF RATIONALITY

A theory of rationality can be viewed as proposing a standard or a norm, such as the values standard. The theory then claims that its proposed

75 Similar suggestions have been made before. Donald Hubin has suggested an account of reasons for action based on what a person “intrinsically values,” and he suggests that our values are “expressions of our selves.” See Hubin 1996, 47. Korsgaard has said that an adequate account of normativity “must appeal, in a deep way, to our sense of who we are, to our sense of identity” (1996, 17–18). For Korsgaard, your identity is “a description under which you value yourself” (101).

76 See the discussion of giving up an ‘ideal’ in Buss 2004, 186–187.

standard is the standard of rationality – the standard that rational agents would comply with to the extent that they can, given their knowledge, just in virtue of being rational. This claim needs to be substantiated by providing the standard with an appropriate ‘grounding.’ Let me explain what would be involved in doing this.

We can formulate any number of purely arbitrary standards. There is, for example, a standard calling on everyone to do a pirouette every night at midnight. We do not think that this standard corresponds to a normative requirement or that it has any bearing on how we are to act. To ground the values standard, I need to show that it is not similarly arbitrary. I need to show that it has a status in virtue of which it is relevantly ‘authoritative’ or ‘normative.’

To show this, I need to show, at least, that the values standard has a status such that those who fail to comply with it have thereby failed in a significant way. There is a problem specifying the *kind* of significance this failure would have. There are irrational people, so we cannot insist that it must be a kind of failure that *anyone* would be motivated to avoid. If the values standard is the standard of rationality, then a knowing failure to serve one’s values would be *irrational*, but to say this would be question-begging and unhelpful. Given the content of the values standard, those who fail to comply with it face a loss, and perhaps a significant loss, since they fail to serve their values as well as they could. This is a kind of loss virtually anyone would be motivated to avoid, but would they be motivated in the way, or for the reason, we are motivated to avoid irrationality?

Donald Hubin suggests that a norm calling on us to pursue our values stands in no need of grounding.⁷⁷ He compares his view of the status of this norm to H. L. A. Hart’s view of the status of the fundamental “rule of recognition” in a legal system. He seems to think that just as, for Hart, it is enough that the rule of recognition be treated a certain way by officials in the legal system, so it is enough that the norm calling on us to pursue our values be treated a certain way by us. At root there are simply certain brute facts, such as that we pursue what we value, and we at least sometimes assess our actions in terms of our values. And Hubin writes: “[T]he property of being rationally advisable just is the property of being properly related to these brute facts.”⁷⁸ However, if the pursuit of what one values is “rationally advisable,” then if a person fails to pursue what she values, this is not simply a departure from a brute psychological

77 Hubin 2001.

78 Ibid., 467; see 463–468. See Hart 1961.

regularity. It is a failure in some interesting normative sense. The challenge is to explain this.

There is an analogous issue about politeness. Consider a rule that calls on us not to wear hats indoors. We might think that this rule is pointless and deny that it has any bearing on how to act, even if we understand that compliance with this rule is locally taken to be a matter of politeness. Nevertheless, in a culture where there are rules that are taken to define politeness, there will be a widely accepted second-order standard, which we could call the “standard of politeness,” that calls on people to comply with the local standards of conventionally acceptable behavior, such as the rule about hats. We cannot suppose that the standard of politeness has a bearing on how we are to act if we think that it is just as arbitrary and pointless as the rule about midnight pirouettes. Those of us who think that politeness has a bearing on how to act – that it is a significant failure of some kind to be impolite – must therefore think that the standard of politeness has some relevant authority or status.

There is an obvious way to think about this. The point of etiquette is to contribute to comfortable and pleasing social interaction. The standard of politeness is relevantly authoritative in virtue of the fact that compliance with it helps to make for comfortable and pleasing social interaction. It plainly is not the case that everyone values comfortable and pleasing social interaction, but the proposal is not that the standard of politeness is authoritative in virtue of the fact that we *value* comfortable and pleasing social interaction. The proposal is instead that there is reason to comply with the standard of politeness in virtue of its status as facilitating social interaction. Call this the “social interaction theory” of politeness.

A complete theory of rationality must be supplemented by a theory that has a similar form and a similar purpose to the social interaction theory of politeness. On anyone’s view, rationality requires acting in certain ways, ways that could be expressed in a standard calling for us to do such and such in such and such circumstances. Any proposal about the content of the standard of rationality is incomplete unless it is accompanied by some account of the basis or authority of the standard.⁷⁹ This, then, is the

79 Elsewhere, I have proposed a schema according to which a basic normative proposition of kind K is true only if a corresponding K standard or norm has a relevant kind of status or justification. See Copp 1995, ch. 2. An account of the grounding of the standard of rationality could be plugged into this schema to yield an account of the truth conditions of claims about rationality. For further discussion of this idea, see chapter 8. See also the [introduction](#) to this book.

grounding problem. To solve it, I think we need to understand the *point* of evaluating people and their actions as rational or irrational.

It might seem that if we can provide an analysis of the *concept* of rationality, and show that a given standard best captures our concept, we have done all that can be done to ground it. I hope that we can do something different from this. Suppose that the concept of rationality is the concept of the kind of virtuosity in the pursuit of what one values specified by the values standard. If so, then we rightly call people who comply with the standard ‘rational.’ But this does not yet give us any reason to view the standard as authoritative. Even if we rightly consider ourselves rational insofar as we conform to the standard, we might wonder whether the standard has any status in virtue of which it actually imposes a requirement on us. Perhaps it will be replied that our concept of rationality is the concept of an *authoritative* standard – so that if the values standard *is* the standard of rationality, it *follows* that it is authoritative. But if this is correct, then if we are in doubt as to whether the values standard is authoritative, we are committed to being equally in doubt as to whether it is the standard of rationality.⁸⁰

The issue is one that Kant apparently had in mind in the *Groundwork of the Metaphysics of Morals*, where he discussed the “possibility” of an “imperative.” In my terms, a Kantian imperative is an ‘authoritative’ standard, one that sets out an actual requirement. The possibility problem for Kant is to explain how to conceive of “the necessitation of the will expressed by an imperative in setting a task” and to show that there is such necessitation.⁸¹ In my terms, that is, regarding a proposed standard of rationality, the problem is to explain how such a standard can be ‘authoritative’ or ‘normative,’ such that it actually imposes a requirement.

10. GROUNDING THE VALUES STANDARD IN AUTONOMY

I have been arguing that the concept of rationality is closely related to the concept of autonomy. To be autonomous is to be *self*-governing, and to

80 Kant writes: “Whoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his [ends] and that lie in his power. This proposition, as far as willing is concerned, is analytic.” Kant 1785, 27 (Ak 417). However, unless there is a requirement of some sort that we be fully rational, then even if it is analytic that fully rational agents do intend the means to their ends, it does not follow that there is a requirement to intend the means to our ends. Hence, in Kant’s terms, it does not follow that the standard calling on us to intend the means to our ends is an *imperative*.

81 Kant 1785, 27 (Ak 417). I discuss some of Kant’s arguments in Copp 1992.

be rational is to do *well* in governing oneself. From one side, this suggests that the *content* of the standard of rationality should be determined by investigating what is involved in governing oneself. From the other side, it suggests a strategy for explaining the *grounding* of the standard of rationality.

The idea in outline is that rationality serves self-government – understood in the thick sense in which self-government involves regulating our actions by our values. The point of assessing the rationality of a person's actions is to appraise her success in manifesting or securing her self-government. That is, if the values standard is the standard of rationality, its grounding or warrant is that to comply with it furthers or instantiates one's self-government. Irrationality is a failure of self-government, a departure from governing oneself well. (Of course, irrational people may be autonomous in the thinner sense; they do poorly at governing themselves in the thicker sense.) The proposal is not that we are rational to manifest and secure our autonomy because we value autonomy. It is that we are rational to manifest and secure our autonomy, even if we do not value autonomy, because manifesting and securing one's autonomy is what rationality consists in.

Call the condition that a standard meets when complying with it furthers or instantiates one's self-government the “autonomy condition.” If what I have argued is correct, it should be plain that the values standard meets this condition. It calls on us to govern ourselves on the basis of our values, or, where our values do not dictate what to do, on the basis of our other intrinsic goals. To act in a way that would not serve our values as well as something else we could do would be to fail to control our actions on the basis of our values. But our values are policies we have that are central to our identity and that are intrinsic in that they are not grounded in any other policies or intentions. They are aspects of our self-conception. Thus, to fail to govern our actions on the basis of our values would be to fail in *self-government*. And in a case where more than one thing we could do would serve our values equally well, to fail to serve our other intrinsic ends as well as we could, while serving our values as well as we can, would be to fail to follow our own policies, despite being able to. This too would be to fail in self-government. Hence, it is plausible that compliance with the values standard furthers or instantiates *governing* oneself.

The autonomy condition is meant to play the same kind of role in grounding the values standard as the condition of furthering pleasing and comfortable social interaction played in grounding the standard of politeness. One might object that to establish genuine normativity, something

much more ambitious than this must be done in order to ground the values standard. For if the grounding of the values standard is analogous to the grounding I proposed for the standard of politeness, then rationality is not a practical virtue that is superior to politeness. There are several things to say in reply to this objection, but I limit myself to two. First, recall that I am working here on the assumption that reasons pluralism is true. I think that there are reasons of etiquette just as there are self-grounded reasons. For this reason, I think it is an advantage that my grounding of the values standard is analogous to the grounding of the standard of politeness. Second, although I accept reasons pluralism, I agree that rationality is a special kind of practical virtue. It is special, not *metaphysically*, but because of its role in *deliberation*, as I will explain.

One might object that since many people and many cultures do not value autonomy, the autonomy condition cannot do the work I want it to do in explaining the authority of the standard of reason. For, one might think, a plausible account of rationality must be culturally neutral and neutral among our substantive values. Moreover, if some of us do not value autonomy, then the fact that compliance with my proposed standard of rationality would serve our self-government would not show that every person has reason to care about being rational.

In response, I want to say, first, that my account does not imply that it is irrational to fail to value autonomy. In fact, it implies (although with a qualification I am ignoring) that it is irrational to fail to promote satisfaction of one's values, *whatever* they are, even if one values a non-autonomous way of life.⁸² According to the values standard, no one is ever rationally required to promote her autonomy unless her own values require her to do so. Second, no plausible theory of rationality can avoid implying that a person might be rationally required to conform with a standard she does not accept. For *any* theory needs to leave room for irrational persons. Finally, virtually anyone *would* want to do what she is required to do, according to the values standard. For the standard merely calls on us to serve our values. When we value something, we naturally are motivated to act appropriately, barring fatigue or depression or the like.⁸³ Motivation by our values is psychologically deep, for we tend to feel shame or guilt or disappointment in ourselves if we fail to act according to our values. Thus, to understand how we can be motivated to act in

82 I refer to the qualification about meeting one's needs. See note 61 above.

83 Valuing something involves having a policy, or a general intention. One who has such an intention has at least a background standing disposition to act appropriately.

accord with the values standard, it is not necessary to suppose that we value promoting our own autonomy in addition to valuing such things as friendship. If irrationality is the failure of self-government, we can see why people typically care to avoid irrationality.

The objection seems to assume that the autonomy condition seeks to justify rationality as instrumental to autonomy. But this is a mistake. The view is, rather, that in being rational we *instantiate* our autonomy. The view is not that self-grounded reasons bind us only insofar as we *desire* to be autonomous, or only insofar as we *value* autonomy.⁸⁴ Many of us do not value being autonomous and have no desire to be autonomous.⁸⁵ Nor does my view depend on the idea that autonomy is *valuable* – though I do not deny that it is. To a first approximation, rationality requires us to serve our *values* because serving our values instantiates being autonomous, and because rationality consists in instantiating autonomy. This is why, *whatever* a person values, she is rationally required to serve her values. My claim is not about what rationality requires her to value. It is about what rationality requires her to do *given* what she values.

11. DELIBERATIVE PRIORITY

My account assigns priority to rationality by comparison with other practical virtues, but not by assigning it a special *metaphysical* status. Instead, it assigns rationality a priority in *deliberation*. When agents with the necessary self-understanding and necessary information deliberate about what to do, the ‘default’ is that their decision is in accord with the values standard.⁸⁶ Of course, in my view an agent who decides otherwise is to that

84 If this were my view, then the argument would show, at best, that the requirement to comply with the values standard is hypothetical, or conditional on our valuing autonomy. This is not so, however. When we are rationally required to do something A, in most cases the requirement is conditional on, roughly, our having values such that doing A best serves those values in the circumstances. But the requirement to comply with the values standard itself is not conditional on our having any values in particular. Whatever our values, in complying with the values standard, we best serve those values.

85 David Velleman argues to the contrary that agency has a “constitutive goal” such that any agent must aim to be in conscious control of her behavior – which he says amounts to the aim to be autonomous. It is in virtue of this fact that agents are subject to reasons. See Velleman 1996, 719. Cf. Velleman 2000 (see note 13, above). My account does not require postulating a “constitutive aim” of agency.

86 Korsgaard holds the stronger view that compliance with the “instrumental principle” is *constitutive of deliberate action*. She writes: “[I]f you don’t put one foot in front of the other you will not be walking and you will get nowhere . . . The instrumental principle is, in this way, a constitutive norm of willing, of deliberate action. If you are going to act at all, then

extent less than fully rational. But the priority-in-deliberation thesis is that it is a law-like truth that, given the nature of practical deliberation, when agents who have approximately accurate beliefs about what they value, and about how to achieve what they value, deliberate about what to do and reach a decision based on their deliberation, other things being equal, they decide to do what they are required to do according to the values standard.⁸⁷ This is the default case.

This thesis is supported by claims I have made about what is involved in valuing something, given that the values standard calls on us to serve our actual values. First, values are policies or general intentions. We would not count you as valuing safety, for example, unless you had a tendency to act with caution when faced with known dangers. Second, your values are an aspect of your identity. This means that emotions of self-esteem are harnessed to your values and help to ensure that you have a tendency to pursue what you value. Moreover, the thesis is supported by a picture of the nature of practical deliberation – deliberation that leads to decisions about what to do. According to this picture, in central cases, practical reasoning involves reasoning from general intentions or policies to specific intentions in action. It is means–end reasoning in the sense that it concerns how to carry out general intentions. Given this picture, if a person is rationally required to A according to the values standard, then, other things being equal, she would form the intention to A if she were to deliberate cogently about what to do in light of her values.

There can be exceptions. The person might be depressed or exhausted and might for this reason fail to form any intention to act. Or she might fail to reach any conclusion about how best to serve her values. She might be unable to see how to resolve conflicts among her values. She might not understand what she values, or she might lack relevant information about how to serve her values. Deliberation might lead her to become perplexed about what she values. She might lack the fortitude to resist an impulse to act contrary to her values. She might be *akratic*. Or, perhaps paradoxically, she might *have* the strength to resist the impulse to act in *accord* with her

you must conform to it” (1997, 249). But Korsgaard insists that failures to conform to the instrumental principle must be possible if the principle is normative (247–248, 228). Surely, however, a person who fails to conform to the instrumental principle may well have acted. Agents sometimes perform *akratic* actions. If the climbers had failed to serve their values in the Everest example, but had instead stopped to help the victims, their helping would have been an action. It therefore seems implausible that conformity with the instrumental principle is constitutive of “deliberative action.”

87 The idea of a defeasible law-like generalization is discussed in Lance and Little 2004.

values. In a moment of clarity, she might see that her values are morally unacceptable and decide to do the right thing. The climbers in the Everest example might decide to abort the climb and to help the victims, thereby acting out-of-character and irrationally, but doing the right thing.

With these caveats understood, we can see that the account of rational agency I have proposed explains how reasoning about what to do can lead to a decision, and thus to the forming of an intention. And we can see how rational decision-making can be the default case, assuming approximately true relevant beliefs. Our values are partly constituted by general intentions or policies, and the default is to be guided by them to form specific intentions that will implement, further, or express our values. Suppose you are trying to decide whether to watch a tennis match or to read a novel. Your decision likely will turn on such considerations as how much you are enjoying the novel, how often you find time to watch tennis or to read, who is playing in the match, and so on. Let us suppose that you value the simple pleasures of life, which, for you, include reading novels and following tennis. If you realize that this is the only thing you value that will be affected by your choice, you will pay attention to which of the activities in question promises more enjoyment. Once you have reached a conclusion about this, if all goes well, your value – which is, *inter alia*, a general intention – will lead you to form a specific intention either to read or to watch tennis. In the default case, if you have been reasoning with accurate information, your intention will be to do what the values standard implies that you rationally ought to do.

The important point here is that self-grounded reasons – facts about the impact the agent's alternatives would have on the things she values – have a role in deliberation that other kinds of reasons do not generally have. When agents who have the necessary self-understanding and the necessary information decide to act, the default is that their decision is in accord with the values standard. The corresponding thing cannot be said of morality or etiquette, for example. It is not the case that when a person with the necessary self-understanding and information decides what to do, the default is that she decides to act morally or politely. This is true only of people with the corresponding values.

One might object that a person who is deliberating about what to do is trying to decide what to do *period*, not merely to decide what *rationally* to do. To see this, consider again the Everest example. Given the values of the climbers, suppose that the values standard would require them to press on to the summit. This would mean that carrying on is what the climbers ought *rationally* to do. But I have been supposing that they ought morally

to stop and help the victims. One might object that a reflective climber would want to decide what to do *period*, not merely what she ought rationally to do. On my own view, it might seem, the values standard cannot tell the climbers what they ought to do *period*. For as I said earlier, I hold that there is no answer to the question of what the climbers ought *simpliciter* to do in the imagined situation. That is, I am assuming that there *is* an answer to the question of what they ought rationally to do, and that there *is* an answer to the question of what they ought morally to do, but I hold that there is no overarching normative standard that determines what they ought to do *simpliciter*. Now it is not part of my goal in this chapter to defend my view about ‘ought *simpliciter*,’ so I could avoid the objection by giving up my view. I do not need to do this, however; for the objection is based on a misunderstanding.

The upshot of practical deliberation is not a belief about what one ought to do; it is a decision or an intention. On my view, the values standard cannot tell the climbers what they ought to do *simpliciter*, but only what they ought *rationally* to do. But the climbers are trying to decide what to do. They are not trying to decide what to believe to be the rational thing to do. And a climber’s decision about what to do would be a flat-out decision, a decision either to help the victims or to carry on with the climb. I agree that a reflective climber would want to decide what to do *period*, but on my account she would do so. That is the nature of decisions.

It is true that if someone decides to do something, and if her deliberation is relevantly informed and fully rational, then she ought *rationally* to do it. This does not entail, however, that she ought *simpliciter* to do it. For deliberation is carried out from one’s own practical standpoint, not from the standpoint of a metaphysically overarching standard that determines what one ought to do *simpliciter*. I have argued, in effect, although with caveats, that one’s practical standpoint is constituted by one’s values. If a climber decides to carry on with the climb, she would be confused if, after the climb, she claimed to have made the mistake of failing to decide what to do *period*. Her decision may have been morally indefensible, but it was a flat-out decision. Her mistake was a moral mistake, not the mistake of failing to make a flat-out decision.

12. CONCLUSION

I have distinguished two issues that a theory of rationality must address. First, it must specify the content of the standard of rationality. Second, it must ground the standard in order to support its claim that the standard

is the standard of rationality. I have proposed the values standard, and I have claimed to ground the values standard in the autonomy condition. *Rationality* is in the service of self-government, I have argued.

In closing, I would like to emphasize the modesty of the view. According to the autonomy conception, rational agents are necessarily disposed to comply with self-grounded reasons, just in virtue of being rational. But rational agents are not necessarily disposed to comply with moral reasons, even if they recognize that there are moral reasons that bear on their actions. A rational person *could* have moral values, so moral reasons *could* be reasons for which she acts. A morally virtuous agent has exactly the moral values it is morally best to have, and thus, for such a person, moral reasons are also self-grounded reasons. It is not necessarily true, however, that rational agents have moral values, so it is not necessarily true that they are rationally required to do what they believe they have moral reason to do. A person who ignores reasons of this kind is not necessarily failing to govern herself in light of her values.

REFERENCES

- Appiah, K. Anthony. 1994. "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction." In Charles Taylor et al., *Multiculturalism: Examining the Politics of Recognition*, ed. Amy Gutman, 149–164. Princeton, N.J.: Princeton University Press.
- Aristotle. 1998. *Nicomachean Ethics*. Trans. David Ross, J. O. Urmson, and J. L. Ackrill. Oxford: Oxford University Press.
- Arpaly, Nomy. 2000. "On Acting Rationally Against One's Best Judgment." *Ethics* 110: 488–513.
- Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- . 1996. "Identification, Decision, and Treating as a Reason." *Philosophical Topics* 24: 1–18.
- . 2000. "Reflection, Planning, and Temporally Extended Agency." *Philosophical Review* 109: 35–61.
- . 2003. "Autonomy and Hierarchy." *Social Philosophy and Policy* 20: 156–176.
- . 2004. "Planning Agency, Autonomous Agency." In *Personal Autonomy*, ed. James Stacey Taylor, 35–57. Cambridge: Cambridge University Press.
- Broome, John. 1999. "Normative Requirements." *Ratio* [new series] 12: 398–419.
- Buss, Sarah. 2004. "The Irrationality of Unhappiness and the Paradox of Despair." *Journal of Philosophy* 101: 167–196.
- Cohon, Rachel. 1993. "Internalism about Reasons for Action." *Pacific Philosophical Quarterly* 74: 265–288.
- Copp, David. 1992. "The 'Possibility' of a Categorical Imperative: Kant's Groundwork, Part III." *Philosophical Perspectives* 6: 261–284.

- . 1993. "Reason and Needs." In Frey and Morris 1993: 112–137.
- . 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.
- . 2001. "Rationality, Autonomy, and Basic Needs." In *Being Humans*, ed. Neil Roughley, 334–355. Berlin: de Gruyter.
- . 2002. "Social Unity and the Identity of Persons." *Journal of Political Philosophy* 10: 365–391.
- Darwall, Stephen. 2002. *Welfare and Rational Care*. Princeton, N.J.: Princeton University Press.
- Deigh, John. 1995. "Shame and Self-Esteem: A Critique." In Dillon 1995: 133–156.
- Dillon, Robin S., ed. 1995. *Dignity, Character, and Self-Respect*. New York: Routledge.
- Fischer, John Martin, and Mark Ravizza, S. J. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Foot, Philippa. 1978a. "Moral Beliefs." In Foot 1978c: 110–131.
- . 1978b. "Morality as a System of Hypothetical Imperatives." In Foot 1978c: 157–174.
- . 1978c. *Virtues and Vices*. Berkeley: University of California Press.
- Frey, R. G., and Christopher Morris, ed. 1993. *Value, Welfare, and Morality*. Cambridge: Cambridge University Press.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Greenspan, Patricia. 1975. "Conditional Obligations and Hypothetical Imperatives." *Journal of Philosophy* 72: 259–276.
- Hart, H. L. A. 1961. *The Concept of Law*. New York: Oxford University Press.
- Harman, Gilbert. 1993. "Desired Desires." In Frey and Morris 1993: 138–157.
- Hubin, Donald C. 1991. "Irrational Desires." *Philosophical Studies* 62: 23–44.
- . 1996. "Hypothetical Motivation." *Noûs* 30: 31–54.
- . 2001. "The Groundless Normativity of Instrumental Rationality." *Journal of Philosophy* 98: 445–468.
- Jackson, Frank. 2003. "Cognitivism, A Priori Deduction, and Moore." *Ethics* 113: 557–575.
- Kant, Immanuel. 1785. *Grounding of the Metaphysics of Morals*. Trans. James W. Ellington. Rpt., Indianapolis, Ind.: Hackett, 1981.
- Kitcher, Philip. 1992. "The Naturalists Return." *Philosophical Review* 101: 53–114.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge: Cambridge University Press.
- . 1997. "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut, 215–254. Oxford: Clarendon Press.
- Krakauer, Jon. 1998. *Into Thin Air: A Personal Account of the Mount Everest Disaster*. New York: Anchor.
- Kymlicka, Will. 1995. *Multicultural Citizenship*. Oxford: Oxford University Press.
- Lance, Mark, and Margaret Little. 2004. "Defeasibility and the Normative Grasp of Context." *Erkenntnis* 61: 435–455.
- Lawrence, Gavin. 2002. "The Rationality of Morality." In *Virtues and Reasons*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, 89–148. Oxford: Clarendon Press.
- Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.

- Oshana, Marina. 1998. "Personal Autonomy and Society." *Journal of Social Philosophy* 29: 81–102.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Pettit, Philip. 1991. "Consequentialism." In *A Companion to Ethics*, ed. Peter Singer, 230–240. Oxford: Blackwell.
- Pettit, Philip, and Michael Smith. 1990. "Backgrounding Desire." *Philosophical Review* 99: 565–592.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Schmidtz, David. 1995. *Rational Choice and Moral Agency*. Princeton, NJ.: Princeton University Press.
- Searle, John. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Taylor, Gabriele. 1995. "Shame, Integrity, and Self-Respect." In Dillon 1995: 157–179.
- Velleman, J. David. 1996. "The Possibility of Practical Reason." *Ethics* 106: 694–726.
- . 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Wallace, R. Jay. 2001. "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint* 1: 3–10. [Http://www.philosophersimprint.org/001003](http://www.philosophersimprint.org/001003).

Index

- agency, autonomous, 314–321. *See also* autonomy
- akrasia* (weakness of will), 314, 348. *See also* rationality, kinds of failure of
- Alice (counterexample to judgment internalism), 5
- antirealism. *See* error theory; noncognitivism; realism, quasi-
- apartheid example, 322
- Appiah, K. Anthony, 322
- arguments: from concept acquisition, 106; from disagreement, for naturalism, 47–53, 98; expressivist, against realism, 199, 264–266; is/ought, against naturalism, 12–13, 199; from normativity, against naturalism, 199, 251–253, 264; *reductio* against Reason-as-such, 301–303. *See also* Mackie, J. L.; Moral Twin Earth argument; open question argument
- Aristotle example, 330
- Armstrong, David, 38
- Arrow, Kenneth, 304
- assertion, 160, 186; and belief (Grice), 164, 179; distinction between truth condition and sincerity condition in, 164; expressing states of mind, 160; and intentions, 160, 186; moral, 155, 187–188, 253; and sincerity, 164, 186
- Audi, Robert, 46, 94–95, 99–102, 104, 105; and central claim of intuitionism, 101; and indefeasibly justified beliefs, 99; self-evident propositions defined by, 94, 99
- autonomy: external and internal conditions of, 315; and identity, 317–321, 326–328; and rationality, 333; and self-governance, 310, 314, 316–317; ‘thick’ vs. ‘thin’ conceptions of, 315, 316–317; and the values standard, 344
- Ayer, A. J., 161
- Bach, Kent, 169, 173, 175, 176, 180
- belief-in-God example, 305, 306
- beliefs, economic, 73–75
- beliefs, moral, 8, 253; and moral facts, 83–87; and naturalistic truth conditions, 71–75; qualifying as knowledge, 72–75; synthetic, and empirical evidence, 35, 45–47, 59. *See also* epistemology, moral; moral convictions
- Benacerraf, Paul, 66
- best code (Copp), 237, 243. *See also* S-ideal moral code
- Blackburn, Simon, 8, 83, 145, 155–159, 161, 194–195, 265; and quasi-realism, 8, 145, 156, 194, 195; and the realist surface, 145–146
- Bonjour, Laurence, 41
- Boyd, Richard, 4, 41, 135, 203–207, 209–210
- Bratman, Michael, 185, 316, 317–321, 333, 339; autonomous agency, model of, 317–319, 321; intentions, defined by, 185, 316; policies, defined by, 316
- Brink, David, 4, 41
- bullfighting example, 49–50
- cancelability. *See* coloring; implicature, conventional
- “Canuck” example, 170
- capital punishment example, 24, 72, 86, 90, 139, 155–159, 189, 238
- cat example, 59, 75–77, 79
- causal theory of reference, 206. *See also* Boyd, Richard

- Celsius scale example, 23
- Chirac, Jacques, 24
- Churchland, Paul, 88; and prototype theory, 88
- claims, moral, 3, 266. *See also* beliefs, moral
- code, moral, 16–17, 88. *See also* society-centered moral theory
- cognitivism. *See* noncognitivism
- coloring, 168–171, 175–177, 199; alterability, 172–173; cancelable implications of, 171–174, 179, 191; and conventional implicature, 177; decoloring a colored term, 192; detachable implications of, 173–174, 179, 191–192; and misuse of terms, 171, 191, 192; and moral terms, 191–193; of pejorative terms, 168–171; tests for, 173–174, 191–192; truth test for (Frege), 171, 191. *See also* expressivism, realist; implicature
- common sense principles (Copp), 93, 107
- computational limitations (Field), 44
- conative states, 8
- concept acquisition, argument from, 106
- concepts, normative, 252
- conceptual falsehoods, 125. *See also* conceptual truths
- conceptual truths, 121, 122, 128; derivative, 97, 123, 124; and internal truth conditions, 122; and intrinsic truth-ensuring properties, 122–123; strict, 122–123; wide, 123. *See also* epistemology; propositions
- considerations, moral, 261–262, 270. *See also* reasons, moral
- consequentialism, 114, 117, 119, 135, 139
- constructivism, 19–20. *See also* practical theory; society-centered moral theory
- content, semantic, 160, 175–181
- context of decision (Copp), 188. *See also* reasoning, and deliberation
- contingent a priori truth. *See* Kripke, Saul; propositions
- conventions: expressive, 174; linguistic, kinds of, 181; semantic, 233–234
- “cur”/“dog” example (Frege), 168–173, 176, 177, 199
- cursing example, 164, 167, 184, 188–192, 198, 265
- Dancy, Jonathan, 13, 98, 119, 122
- Darwell, Stephen, 39, 261, 337, 339
- decisions, 185, 188–189, 199, 313, 350. *See also* reasoning, and deliberation
- deflationism, 155, 156, 157, 193, 195, 231; and deflationist meaning of “true,” 155
- See also* noncognitivism, sophisticated; realism, quasi-
- Deigh, John, 7, 263
- deliberation, and reasoning, 313, 333, 336, 347–350
- depression example, 267
- descriptivism, 7, 264
- disagreement, 212; and Moral Twin Earth argument, 209–243; and translation, 212–216, 242–243
- discourse: moral, 200; normative, 198
- dollar bill example, 73–75
- Dummett, Michael, 170
- education, moral. *See* moral education
- egoism, ethical, 288
- Ellesmere Island example, 330
- emotions of esteem/self-assessment, 321, 323–326
- empirical indefeasibility, 98, 99
- endorsement, 19, 185, 265–266, 320, 327–328
- epistemology: default reasonable beliefs, 46; of mathematics (Benacerraf), 66; normative propositions of, 15; and practical reasons for belief, 305; and psychology, 61, 65–67; reliabilism, 74–75; and sensitivity to kinds of fact, 75–78. *See also* conceptual truths; epistemology, moral; epistemology, naturalized
- epistemology, moral, 56, 63, 64–65, 89–91; central problem for, 84; and experience, 40–41, 42, 104; and moral sensitivity, 75–76, 77, 78; and relation to moral facts, 85; and society-centered moral theory, 72; suppositions for nonskeptical, 86. *See also* intuitionism; self-evidence thesis
- epistemology, naturalized: central doctrines of, 62; characterizations of, 64; defined, 61–68; and ethical naturalism, 67, 87–91; and scientific prioritism, 67
- error theory (Mackie), 9, 83, 103, 117, 119, 127, 133, 208–228; and the falsity of basic moral propositions, 9; and the Mackie test, 118; two claims of, 117. *See also* nihilism
- etiquette, 14–15, 271, 312, 343, 345. *See also* standard-based account of normative judgment
- euthanasia example, 59–60
- Everest example, 311–313, 314, 338–339, 349–350
- evidence: testimonial, 44; unjustified vs. false, 99
- excellence: Aristotelian, 298; personal, 297–300
- exemplar view, 88. *See also* prototype theory

- expression, 162–164; conversational, 188;
 Frege-, 174–181, 182, 186, 193, 194, 199;
 sincerity, 164–166. *See also* expressivism
- expressivism, anti-realist, 153, 156, 161–162,
 231; argument for, 264–266; distinguished
 from moral realism, 7–8, 156, 157;
 distinguished from realist-expressivism,
 194–197; and moral predicates, 158; and
 moral properties, 158. *See also*
 expressivism, realist; noncognitivism;
 realism, quasi-
- expressivism, realist, 9; and anti-realist
 expressivism, 194–195, 264–266;
 arguments for, 186; and basic moral
 judgments, 162; and conventional
 implicature, 178; and conversational
 implicature, 178; and judgment
 internalism, 193–194; *see also* internalism,
 judgment; and relation between moral
 assertion and conative state of mind,
 162–167; and subscription to a moral
 standard, 184, 186; three kinds of, 178–181
- family example, 332
- famine relief example, 165–167, 178,
 193–194
- Field, Hartry, 42, 43, 44, 51, 97, 105, 106,
 157; and the *a priori*, 42, 97, 106; and
 testimonial evidence, 43
- Foot, Philippa, 261–262, 321
- Frege, Gottlob, 9, 168–171, 190, 191. *See also*
 coloring; “*cur*”/“*dog*” example
- friendship example, 35, 45–47, 98, 328–329
- Gauguin example, 297–298, 299
- Genesis*, 38
- Gibbard, Allan, 39, 83, 159, 161, 164, 182,
 184–185, 198–199, 265–266; on assertion
 and truth and sincerity conditions, 164;
 and norm acceptance, 162, 184; and
 thinking something rational, 162
- Gödel, Kurt, 66
- gold example, 107, 108
- Goldman, Alvin, 88
- good girl example, 195–196
- goodness, moral, 207
- gopper example, 131–132
- Greek ancestry example, 324–325
- Greene, Graham, 259
- Grice, Paul, 171–174, 177, 179, 191; on
 assertion and belief, 179; on implicature,
 173, 177, 179, 191, 230
- Gyges, 262–263, 277–278, 279–280,
 286–288, 289, 292, 295, 296, 298, 299
- H₂O proposition, 128–130. *See also* water
 example
- Hare, R. M., 1, 161, 166–167, 184, 194,
 195–196, 265; on descriptive meaning vs.
 evaluative meaning, 195–196, 197; and
 inverted commas, 166–167, 191; on moral
 judgment, 161
- Harman, Gilbert, 40, 59, 75
- Hart, H. L. A., 342
- Hill, Thomas E., Jr., 169
- homeostatic cluster property (Boyd), 135
- honesty example, 328–329
- honey example, 206
- Horgan, Terence, 27, 203–213, 231–232;
 Moral Twin Earth argument of, 208–228;
 synthetic semantic naturalism, defined by,
 211–212
- Hubin, Donald, 313–314, 338–342
- Hume, David, 120
- Hurka, Thomas, 298
- identity of persons: and autonomy, 317,
 321–328; Lockean account of, 318; and
 the practical standpoint, 321, 323. *See also*
 Bratman, Michael; self-esteem identity
- implicature: and belief, 179; conventional,
 177, 178, 179, 180–181; conversational,
 171, 178, 180, 187–188, 192. *See also*
 coloring
- indeterminacy: semantic, 234–236, 243
- indoor hat example, 343
- intentions: and desires, 315; as planful states
 (Bratman), 185, 316; and policies, 316;
 prior, 316; referential, 221–245; semantic,
 233–238, 243, 245. *See also* Putnamian
 semantics
- internalism: discourse (Copp), 194, 199;
 motivational, 258–260, 267; normative
 (Copp), 4. *See also* judgment, internalism
- intuition, semantic (Horgan and Timmons),
 210
- intuitionism, 46, 57, 66, 87; defined, 46. *See*
also Audi, Robert
- inverted commas sense (Hare), 166–167, 191
- Jackson, Frank, 37, 135–137, 180; and
 mature folk morality, 135, 137; and moral
 functionalism, 135; supervenience defined
 by, 36, 118
- jade example, 218
- jazz example, 331, 332
- Jubien, Michael, 120
- judgment, moral, 1, 160, 161, 291
- judgment, normative. *See* standard-based
 account of normative judgment
- judgment internalism, 1, 5–6, 8, 154–155;
 arguments against, 5–6, 258–259, 260; *de*
dicto vs. *de re* (Tresan), 260; and
 normativity, 6, 258–259, 260

- judgment internalism (*cont.*)
 and realist-expressivism, 199. *See also* noncognitivism
- justice, and the basic structure of society, 21
- Kant, Immanuel, 41, 50, 143–145, 275, 310, 312, 344–347; *Groundwork of the Metaphysics of Morals*, 310, 344–347
- King, Jeffrey C., 40, 55, 129, 182, 253
- knowledge. *See* epistemology
- Kornblith, Hilary, 63
- Korsgaard, Christine, 270, 276–279, 313, 322; and identity of persons, 277, 278; and moral naturalism, 278; and the normative question, 276–277; procedural realism defined by, 277; substantive realism defined by, 276
- Kraut, Richard, 290, 298
- Kripke, Saul, 98, 106, 107, 122
- learning, moral. *See* moral education
- Lewis, David, 136
- Locke, John, 318–319, 320, 321
- lying example, 9, 14, 48, 117, 209–217, 220, 221, 222, 257, 266
- M-necessity, 120
- M-worlds, 113–114, 120
- Mackie, J. L., 9, 83, 117–119, 133, 144, 251; and the argument from queerness, 9, 251; and the property of not-to-be-doneness, 223, 225; Mackie-substantive necessary truths, 118, 126; Mackie test, 118, 123. *See also* error theory
- Matty example, 43, 45
- mature folk morality (Jackson and Pettit), 136
- McNaughton, David, 119
- metaethics, i, 2
- midnight pirouette example, 342
- “milk” example, 204–205, 218–227, 233–234
- Miller, Richard, 41
- miserly tip example, 324
- Moore, G. E., 11, 34–35, 36, 37, 41, 126, 134, 203, 251, 255; natural properties defined by, 34–35; naturalism defined by, 255; paradox of, 163, 179. *See also* open question argument
- moral convictions (Copp), 185, 187, 200
- moral education, 78–83, 84, 88, 186–187, 198, 220, 307
- moral expertise, 82
- moral functionalism (Jackson and Pettit), 135; central objection to (Copp), 137
- moral necessities. *See* necessary truths, moral
- moral progress, 86
- Moral Twin Earth argument (Horgan and Timmons), 108, 203–228, 231, 234, 236–237, 239–245; Master scenario for (Copp), 212–216
- morality: function of, 13; as overriding, 285–286, 301, 302; and self-interest, 284–289, 306–308; as a system of justified standards, 296. *See also* society-centered moral theory
- Mother Theresa example, 37
- mountain climbers example. *See* Everest example
- multiple propositions framework, 176, 177, 178–181; coloring and conventional implicature in, 177
- murder example, 96, 116
- narwhale tusks example, 218–220
- naturalism: and science, 39, 56; and reductionism, 11–12, 133–141, 253; unrestricted, 34. *See also* naturalism, moral
- naturalism, moral, 2, 11, 13–26, 33, 40, 55–56, 68–71, 94, 96–99, 253–255, 268; as an assimilation between ethics and science, 39; best understood as strongly a priori, 45–46, 53; challenges to, 58–68, 71–75; *see also* arguments and open question argument; conceptual, 133; as default view, 2, 10–13; defined (Copp), 55, 67, 230, 249, 254; defined (Moore), 35, 55, 255; doctrines characteristic of, 10–13, 94; epistemic construal of (Shafer-Landau), 95; metaphysical, 133; and moral necessities, 132–141; *see also* necessary truths, moral; and moral properties, 33, 34, 251–253; and natural properties, 203; and normativity of moral judgment, 33, 252, 264; synthetic, 204–208, 211–212, 230; and T-identity theses, 11–12, 133–141. *See also* Moore, G. E.; moral functionalism; realism, moral, Cornell; society-centered moral theory
- naturalistic fallacy, 199
- Neale, Stephen, 175–176, 177. *See also* multiple propositions framework
- necessary truths, 121; moral, 113–121, 148; trivial, 116. *See also* conceptual truths; necessary truths, moral, substantive; possible worlds
- necessary truths, moral, substantive, 114, 124; as conceptual truths, 124–128; as metaphysically necessary, 128–141; and moral naturalism, 132–141; moral necessities constraint defined, 114, 147; and noncognitivism, 145–147; as practical necessities, 142–145
- nihilism, 2, 120. *See also* error theory
- noncognitivism, 2, 8, 143, 145, 208; and antirealism, 143–145; argument from judgment internalism for, 8; and deflationism, 7; and moral necessities, 146;

- negative doctrine of, 145–147; positive doctrine of, 145; sophisticated versions and defensive doctrine of, 146–147. *See also* expressivism, antirealist
- nonnaturalism, moral, 133
- norm acceptance (Gibbard), 162, 184–185
- normative question (Korsgaard), 276–277
- normative standpoints, 285, 294, 296, 307–308. *See also* overridingness; ought *simpliciter*; Reason-as-such; reasons, kinds of
- normativity, 1, 2, 252–253; generic, 255–258, 266–269, 274; grades of, 27, 255, 275–280; as internal to moral judgment, 6; motivational, 256–257, 258–260, 264, 266–269, 271–274; and normative properties, 252–253; and reasons, 256, 285; as second-order property of moral properties, 252. *See also* normativity, authoritative; normativity, constraint
- normativity, authoritative, 256–257, 260–263, 270–271, 274–275, 281; against, 280–281; authoritative reasons proposal, 261, 262, 270, 271–272; closed question proposal, 262, 270, 272; Kantian approaches to, 275; and naturalism, 274–275; and moral reasons, 274
- normativity, constraint, 2, 3–6
- norms, 182, 290–291, 296
- not-to-be-doneness (Mackie), 223, 225
- obsession example, 327
- open question argument (G. E. Moore), 11–13, 126, 134, 199, 203, 244, 251
- order, causal, and the natural world, 37–38
- ought *simpliciter*, 284–285, 286, 292–296, 299, 302–303, 313, 350. *See also* overridingness; Reason-as-such
- overridingness, 284, 292–308. *See also* ought, *simpliciter*; normativity, authoritative; Reason-as-such
- pain proposition (Copp), 103–104
- painful death proposition (Copp), 116–120, 124–128, 130–132, 134, 139–141, 148
- pantheism, 38
- particularism, 48, 119
- Pettit, Philip, 135–137, 180
- Plato, 262, 286–287
- policies, 316, 328; and intentions, 316, 333; intrinsic, 329; quasi-policies (Bratman), 319; second-order, 318
- politeness example, 343
- possible worlds, 120
- practical theory (Copp), 20, 21. *See also* constructivism
- pragmatics, 155
- predicates: moral, 156–158, 235, 253, 264; property-ascribing, 231
- prescriptivity, 1
- principles, moral, 109, 113–114; sadist's, 109
- priority, deliberative. *See under* rationality, self-grounded
- properties: constituents of, 131; disjunctive, 131–132; normative, 290. *See also* properties, moral; properties, natural
- properties, moral, 7, 253, 265; according to nonnaturalism, 65; as empirical, 35, 254; metaphysically robust, 157; as natural properties, 43, 70–71, 254; as normative, 4, 10, 252, 266; psychological role of, in shaping action, 10; as relational, 23–25, 110. *See also* society-centered moral theory
- properties, natural, 254; contemporary conceptions of, 36–39; defined (Copp), 11, 39, 43, 96, 97, 98, 133, 254; defined (G. E. Moore), 34–35; defined (Shafer-Landau), 96; as empirical, 39–41, 55–56; epistemological characterizations of, 39; metaphysical characterizations of, 37–38; ostensive definitions of, 37; reductionist and relational approaches to, 36–37. *See also* naturalism, moral
- propositions, 160; about the instantiation of a property, 39, 56; a priori, 41–42; empirical, 41–42; grasping of, 102, 121, 129. *See also* conceptual truths; propositions, moral; propositions, normative; propositions, self-evident; propositions, synthetic
- propositions, moral; basic, 9, 68, 110, 138, 157, 182, 183, 254; distinguished from moral standards, 14, 15, 137; and moral terms, 182; pure, 138. *See also* standard-based account of normative judgment
- propositions, normative, 137–138, 290; basic, 14, 138; kinds of, 15–16; pure, 14, 138. *See also* standard-based account of normative judgment
- propositions, self-evident, 93, 99–102, 128; and adequate understanding and attentive consideration, 100–102, 103, 104, 105, 128; and a priori knowledge, 94–95, 104–106; definitions of, 93, 94, 99, 100–102, 109; empirical defeasibility of, 110–111; relativized to believer's cognitive situation, 109. *See also* self-evidence thesis
- propositions, synthetic, 40, 56, 97, 98, 130; a priori, 41–45; a priori strong, 42–53, 97; a priori weak, 42, 46, 97

- prototype theory, 88. *See also* exemplar view
- Putnam, Hilary, 135, 203–209, 217–228, 232–245; Twin Earth thought experiment of, 208, 209, 215. *See also* Putnamian semantics
- Putnam–relation, 206
- Putnamian semantics, 204–205, 206, 217–220, 232–245; fallback conditions in, 217, 222; and moral terms, 207; and Moral Twin Earth argument, 238–243; and mistakes, 219; ostensive definitions in, 217–223; referential intentions, 204, 219–245; and semantic indeterminacy, 243; and society-centered theory, 238
- quasi-realism (Blackburn). *See* realism, quasi-.
See also expressivism, antirealist; noncognitivism
- Quine, W. V. O., 41, 61–62, 63–64; on epistemology and psychology, 61; “Epistemology Naturalized,” 61–62
- Railton, Peter, 4, 39, 41
- rain example, 176
- rationalism: moral, 295–296; self-grounded, 295, 296. *See also* overridingness; Reason-as-such
- rationality, practical, 290; autonomy conception of, 309, 311–314, 351; ungrounded (Hubin), 313–314, 342; grounding a conception of, 313, 341–347; *see also* standard-based account of normative judgment, and truth-grounding status; kinds of failure of, 314, 336, 348; and moral virtue, 312, 351; as a practical virtue, 347; requirements of, 334; and self-governance, 310, 333; and wisdom. *See also* rationality, self-grounded
- rationality, self-grounded (Copp), 255, 273–280, 294–297, 299, 302, 307–308, 310–351; deliberative priority of, 313, 347–350; instantiating autonomy, 347; and needs, 335; objection from ungrounded ends, 337–341; and values standard, 310, 313, 334, 344; and values, 333–337. *See also* rationality, practical
- Rawls, John, 21, 80
- realism, moral, 7, 156, 267; central doctrine of, 195; Cornell, 4–5; as default view, 6–9; as descriptivism, 7; distinguished from antirealist expressivism, 7–8, 153, 155–158; five doctrines included in, 7, 253; internalist, 154; and Kant, 143. *See also* expressivism, anti-realist; expressivism, realist
- realism, quasi-, 8, 145, 156, 194, 195
- realist-expressivism. *See* expressivism, realist
- reason, practical; Hobbesian, 288; instrumental, principle of, 309; Kantian conception of, 288, 310; unity of, skepticism about, 300–306. *See also* rationality, practical
- Reason-as-such, 293–294, 297, 301; comprehensiveness of, 294, 299; incoherence of the concept of, 301–303; and self-grounded reason, 294–297; supremacy of, 299–300, 301–302. *See also* normativity; ought *simpliciter*; overridingness
- reasoning, 84; and deliberation, 313, 333, 336, 347–350; moral, 58–59, 78–81, 82, 83
- reasons: authoritative, 261, 269, 271–272; epistemic, 305–306; of etiquette, 271; genuine, 313; kinds of, 285, 288, 290, 301; moral, 256, 271, 274; and normativity, 256; of personal excellence, 291, 297–300; pluralism, 306, 311, 312, 346; self-grounded (Copp), 291–292, 307, 311; *see also* rationality, self-grounded; of self-interest, 285, 291
- “redneck” example, 12, 265–266
- relationalism, 244. *See also* society-centered moral theory, relativism of
- relativism. *See* relationalism; society-centered moral theory, relativism of
- rules, 88, 140. *See also* norms; standards
- S-ideal moral code (Copp), 70, 82, 89–91, 110, 183. *See also* best code
- Sayre-McCord, Geoffrey, 207
- Scanlon, T. M., 328, 331
- schmoral requirement, 296
- schmurder, 96, 116–117
- schpreet example, 260
- Scobie, 259
- seatbelt example, 330, 340
- self-conception strategy, Kantian, 143–145, 270, 275–279
- self-esteem, 323–327, 328. *See also* emotions of esteem
- self-esteem identity (Copp), 321, 324–328, 341; and autonomy, 326–328; defined, 325; and values, 326–351. *See also* emotions of esteem
- self-evidence thesis, 93–104, 109, 111; and moral naturalism, 105, 106–109, 110. *See also* epistemology, moral; propositions, self-evident
- self-interest, 284–290, 308, 309; and morality, 285, 286–308; objectivist accounts of, 289; subjectivist accounts of, 289
- semantics, 135, 155. *See also* Putnamian semantics

- Shafer-Landau, Russ, 93–95, 96, 99–102, 103, 105, 128
- Sidgwick, Henry, 304
- slave example, 5, 15, 125–126
- Slote, Michael, 298
- Smith, Michael, 12, 134, 165–166, 193–194, 275
- society, 21, 22, 138; basic needs of, 17–18, 69, 237
- society-centered moral theory (Copp), 13–21, 26, 68–69, 71, 110–111, 137, 183–185, 237–238; and basic moral claims, 89–91; basic theory, 19, 21; and constructivism, 18–20; counter-examples to, 140; and moral epistemology, 72; and moral necessities, 137–142; objections to, 25–26, 69–71, 84; original theory, 19–20; and practical rationality, 20; relativism of, 22–25, 244; relevant society in, 23–25, 138, 237; and the T-identity thesis, 139; and the truth-conditions of moral propositions, 14–18, 20, 138. *See also* naturalism, moral; standard-based account of normative judgment
- soda pop example, 204–205
- spatiotemporal manifold (Armstrong), 38
- stances, 8, 145, 161
- standard-based account of normative judgment, 14, 15–16, 182; and reasons, 290; and truth-grounding status, 14, 15–16, 313, 335, 338–342, 343, 344–347. *See also* etiquette; propositions, normative; society-centered moral theory; standards; values standard
- standard meter example (Kripke), 98, 106
- standards, 14, 15–16, 68, 110, 182, 290–291; authoritative or justified, 16, 290, 293–294, 296; epistemic, 335; *see also* epistemology; moral, 14, 69, 87, 110, 137, 183, 307; *see also* standard-based account of normative judgment, and truth-grounding status
- states, conative, 8, 153, 265, 315, 318, 319. *See also* stances
- Stevenson, Charles, 161, 265
- stolen radio example, 324
- Sturgeon, Nicholas, 4, 12, 40
- subscription to a moral standard (Copp), 14, 184; and moral belief, 186–187. *See also* expressivism, realist; moral convictions
- supervenience, 36, 118
- T-identity thesis (Copp), 133–141
- “t-milk” example, 234
- Taylor, Gabriele, 323
- tennis match/novel example, 349
- terms, moral, 195–196, 220; thick and thin, 197; and translation, 216. *See also* coloring thought, moral, 160, 161
- Timmons, Mark. *See* Horgan, Terence
- torture example, 8, 9, 58, 79, 84, 111, 257
- translation, 212–216, 242–243
- Tresan, Jon, 21, 109, 260
- truth. *See* conceptual truths; deflationism; necessary truths; necessary truths, moral, substantive; society-centered moral theory, and the truth-conditions of moral propositions
- tweet example, 259–260
- Twin Earth. *See* Moral Twin Earth argument
- twong example, 223
- understanding, adequate. *See* propositions, self-evident
- unicorn example, 218–220, 221
- values, 328–330; and action, 330, 348; of an autonomous agent, 317–323; and beliefs, 312, 331; first-order, 319, 321; intrinsic, 329, 340; moral, 307, 310–311; as policies for action, 328–333; and practical standpoint, 350; serving, 334. *See also* rationality, practical; rationality, self-grounded; self-esteem identity; values standard
- values standard, 334. *See also* rationality, practical; rationality, self-grounded
- virtuousness, 22
- vixen/fox example, 123
- Wallace, Jay, 327, 339
- Warszawski example, 23
- water example, 203–204, 209–236. *See also* H₂O proposition
- wide reflective equilibrium theory (Rawls), 80
- Williams, Bernard, 297, 298
- Wolf, Susan, 298
- wrongness, 22, 70–71, 103, 107–109, 111, 116–120, 126–128, 130, 132, 136, 146, 198, 207–215, 216–217, 220–225, 267; Millian account of, 15
- “Yankee” example, 170, 174–175, 181, 192